Ecomorphology, biomechanics and ontogeny of the pectoral girdle in anurans (Lissamphibia: Anura) with emphasis on morphological methods

Dissertation

with the aim of achieving a doctoral degree

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Biology

University of Hamburg

submitted by

Karolin Engelkes

Hamburg 2020

Head of examination commission:
Prof. Dr. Susanne Dobler, Universität Hamburg

Dissertation reviewers:
Prof. Dr. Alexander Haas, Universität Hamburg
Prof. Dr. Andreas Schmidt-Rhaesa, Universität Hamburg
Prof. Dr. John A. Nyakatura, Humboldt Universität zu Berlin

Date of oral defense: 19.03.2021

Contents

Abstract

Frogs and toads (Amphibia: Anura) inhabit a wide range of habitats that are each associated with specific modes of locomotion. Despite the generally conserved anuran Bauplan, various anatomical adaptations to locomotor behaviors have been reported in previous studies. Most of these studies, however, focused on the limbs or the pelvic girdle. The pectoral girdle, that is the anatomical complex that connects the forelimbs to the axial skeleton, has rarely been analyzed with regard to its adaptation to and function in locomotion. Various previous studies on the anatomy and evolution of the pectoral girdle skeleton provide a base for ecomorphological and functional analyses of the skeletal girdle elements. The inclusion of the pectoral girdle muscles in such studies is complicated by inconsistencies in the identification and naming of the shoulder joint muscles in literature accounts. Therefore, the aims of this thesis were twofold: The first aim was to analyze the relationships between locomotor mode, skeletal shape variation, and biomechanical function of the anuran pectoral girdle. The second aim was to assess the ontogenetic development and innervation of the shoulder joint muscles in order to resolve the inconsistencies in literature accounts and to thereby provide a base for future studies that could include those muscles in biomechanical analyses or reconstruct muscle evolution.

The first aim was approached by assessing whether geometric morphometrics was a valid method to analyze the shape of the anuran pectoral girdle and by optimizing the corresponding workflow. As geometric morphometrics proved to be a suitable approach, it was used to study the shape diversity in the pectoral girdle bones in relation to locomotor behavior within a phylogenetic framework. The analyses were complemented by musculo-skeletal modelling and finite element analyses in order to understand the biomechanical implications of shape differences in the context of locomotion. Digital dissections of volumes that were generated by histological serial sectioning, episcopic microtomy, or micro-computed tomography of larval and adult specimens were performed to approach the second aim.

Phylogenetic relationships, size, and locomotor behavior had an effect on the shape of the pectoral girdle in anurans, but the relative impact of these factors varied between bones. Remarkable shape diversity was observed within locomotor groups which indicates many-to-one mapping of form onto function. The girdle shapes of burrowing and non-burrowing

species, and headfirst and backward burrowing species significantly differed from one another. The moment arms of (simulated) humerus retractor muscles crossing the shoulder joint were enlarged in burrowing species by specific pectoral girdle geometries. This potential adaptation to burrowing behavior was achieved by different, species-specific mechanisms. Differences in the pectoral girdle shapes were associated with differences in the reaction of the coracoid to simulated loading by physiologically relevant forces.

The anuran shoulder joint muscles were ontogenetically derived from the ventral and the dorsal pre-muscle mass that can be found in all vertebrates. The commonly used names ‘m. coraco-brachialis longus’ and ‘m. deltoideus’ were found to be misleading with regard to the ontogenetic origin of the corresponding muscle units. The mm. scapulohumeralis profundus anterior and posterior, although present in all examined species, have been overlooked in some studies. If present, the portions of the mm. cleidohumeralis, supracoracoideus, and coracobrachialis have occasionally been incorrectly recognized or assigned in previous studies. All other shoulder joint muscles have correctly been identified and named in previous studies. A nomenclature consistent with regard to inter-specific homologies and the ontogenetic origin of muscle units was suggested. Shape variations in the skeletal elements of the pectoral girdle and also of the forelimbs provide the base for various, potentially adaptive configurations of the shoulder joint muscles.

Zusammenfassung

Frösche und Kröten (Amphibia: Anura) leben in vielen verschiedenen Habitaten, in denen sie jeweils spezifische Fortbewegungsweisen verwenden. Trotz des weitestgehend konservierten Bauplans der Anuren haben frühere Studien verschiedene anatomische Anpassungen an die Fortbewegungsweisen gefunden. Die meisten dieser Studien haben sich jedoch auf die Beine und den Beckengürtel konzentriert. Mögliche Anpassungen des Schultergürtel, also des anatomischen Komplexes, der die Vorderbeine mit dem Achsenskelett verbindet, wurden bisher kaum untersucht. Frühere Studien zur Anatomie und Evolution des Schultergürtelskeletts bieten eine Grundlage für weiterführende ökomorphologische und funktionelle Analysen des Skeletts. Die Einbeziehung der Schultergürtelmuskeln in solche Analysen wird jedoch durch die inkonsistente Identifizierung und Benennung der einzelnen Muskeleinheiten in der Literatur erschwert. Darum werden mit dieser Arbeit zwei Ziele verfolgt: Zunächst sollen die Beziehungen zwischen Fortbewegungsweise, Form und Funktion des Schultergürtels der Anuren analysiert werden. Das zweite Ziel besteht darin, die ontogenetische Entwicklung und Innervierung der Muskeln des Schultergelenks zu klären und mit diesen Erkenntnissen die Inkonsistenzen in der Literatur aufzulösen, um damit eine Basis für mögliche zukünftige Studien, die die Schultergürtelmuskulatur in biomechanische Analysen einbeziehen oder die Muskelevolution rekonstruieren könnten, zu schaffen.

Um das erste Ziel zu erreichen, wurde zunächst untersucht, ob die Geometrische Morphometrie eine valide Methode ist, um die Form des Schultergürtels der Anuren zu untersuchen, und der Workflow wurde für diesen speziellen Fall optimiert. Da sich die Geometrische Morphometrie als geeignete Methode herausgestellt hat, wurde mit ihrer Hilfe die Formvarianz der Knochen des Schultergürtels im Kontext von Lokomotion und Phylogenie untersucht. Die Untersuchungen wurden durch Muskel-Skelett-Modellierungen und Finite Elemente Analysen ergänzt, um die biomechanischen Auswirkungen von Formunterschieden zu verstehen. Digitale Präparationen von Volumendaten, die durch histologische Serienschnitte, episkopische Mikrotomie und Mikro-Computer Tomographie gewonnen wurde, wurden an larvalen und adulten Individuen durchgeführt, um das zweite Ziel zu erreichen.

Phylogenetische Verwandtschaft, Größe und Fortbewegungsweise hatten einen signifikanten Effekt auf die Form des Schultergürtels der Anuren; die Stärke des Einflusses

dieser Faktoren unterschied sich jedoch zwischen den einzelnen Knochen. Eine bemerkenswerte Formenvielfalt wurde innerhalb der Lokomotionsgruppen beobachtet, was many-to-one mapping der Form auf die Funktion impliziert. Signifikante Unterschiede in der Form des Schultergürtels wurden zwischen grabenden und nicht-grabenden, sowie zwischen vorwärts- und rückwärts-grabenden Arten beobachtet. Die Hebelarme von (simulierten) Humerus-Retraktormuskeln waren in grabenden Arten durch spezielle Schultergürtel-geometrien vergrößert. Diese mögliche Adaptation an das Graben wurde durch verschiedene, artspezifische Mechanismen erreicht. Die Unterschiede in der Form des Schultergürtels gingen mit Unterschieden in der Reaktion auf simulierte, physiologisch relevante Kräfte einher.

Die Schultergelenksmuskeln der Anuren leiteten sich ontogenetisch von der ventralen und der dorsalen Prä-Muskelmasse, die in allen Vertebraten beobachtet werden kann, ab. Die Ergebnisse zeigten, dass die häufig verwendeten Namen ‘m. coraco-brachialis longus’ und ‘m. deltoideus’ missverständlich in Bezug auf den ontogenetischen Ursprung der zugehörigen Muskeleinheiten sind. Die mm. scapulohumeralis profundus anterior und posterior waren in allen untersuchten Arten vorhanden, wurden in einigen früheren Studien aber übersehene. Die Portionen der mm. cleidohumeralis, supracoracoideus und coracobrachialis wurden, wenn vorhanden, in früheren Studien gelegentlich falsch erkannt oder falschen Muskelgruppen zugeordnet. Die übrigen Schultergelenksmuskeln wurden in der Literatur richtig identifiziert und benannt. Eine Nomenklatur, die hinsichtlich interspezifischer Muskelhomologie und ontogenetischem Ursprung konsistent ist, wurde vorgeschlagen. Die Formvarianz in den Skelettelementen des Schultergürtels und auch in den Knochen des Vorderbeins stellt die Grundlage für verschiedene, möglicherweise adaptive Konfigurationen der Schultergelenksmuskulatur dar.

- Chapter one -

General introduction

With 7243 currently recognized species, frogs and toads (Anura) form the largest of the three orders of extant amphibians (Duellman & Trueb, 1994; Frost, 2020; Fig. 1). In contrast to their sister group, the Caudata (e.g., Pyron & Wiens, 2011; Jetz & Pyron, 2018), anurans are tailless and exhibit elongated hindlimbs (e.g., Duellman & Trueb, 1994). The tibia and the fibula, as well as the radius and the ulna, are fused (Shubin & Jenkins, 1995). The vertebral column of anurans is shortened (5-9 presacral vertebrae), the transvers process of the sacral vertebra are expanded and articulate with rod-like ilia, and the coccyx and hypochord are synostotically fused to form the urostyl (e.g., Duellman & Trueb, 1994; Shubin & Jenkins, 1995; Púgener & Maglia, 2009).

Anurans are distributed across nearly the entire globe and inhabit a wide range of habitats including, for example, ponds, tropical rainforests, or deserts (Duellman & Trueb, 1994; Wells, 2007). Various ecotypes have evolved within those habitats; some species are exclusively aquatic, others are semiaquatic, riparian, terrestrial, fossorial, or arboreal (Wells, 2007). The ecological diversity is accompanied by a diversity in locomotor behaviors: While almost all anuran species are capable of swimming (Abourachid & Green, 1999) and hopping or jumping (defined as leaps with a maximum length of, respectively, more or less than 8-9 times the snout-vent-length; Emerson, 1979; Wells, 2007), some species also use extensive quadrupedal walking (e.g., Ahn et al. 2003). Burrowing and climbing locomotion have both evolved several times independently (e.g., Nomura et al. 2009; Reilly & Jorgensen, 2011).

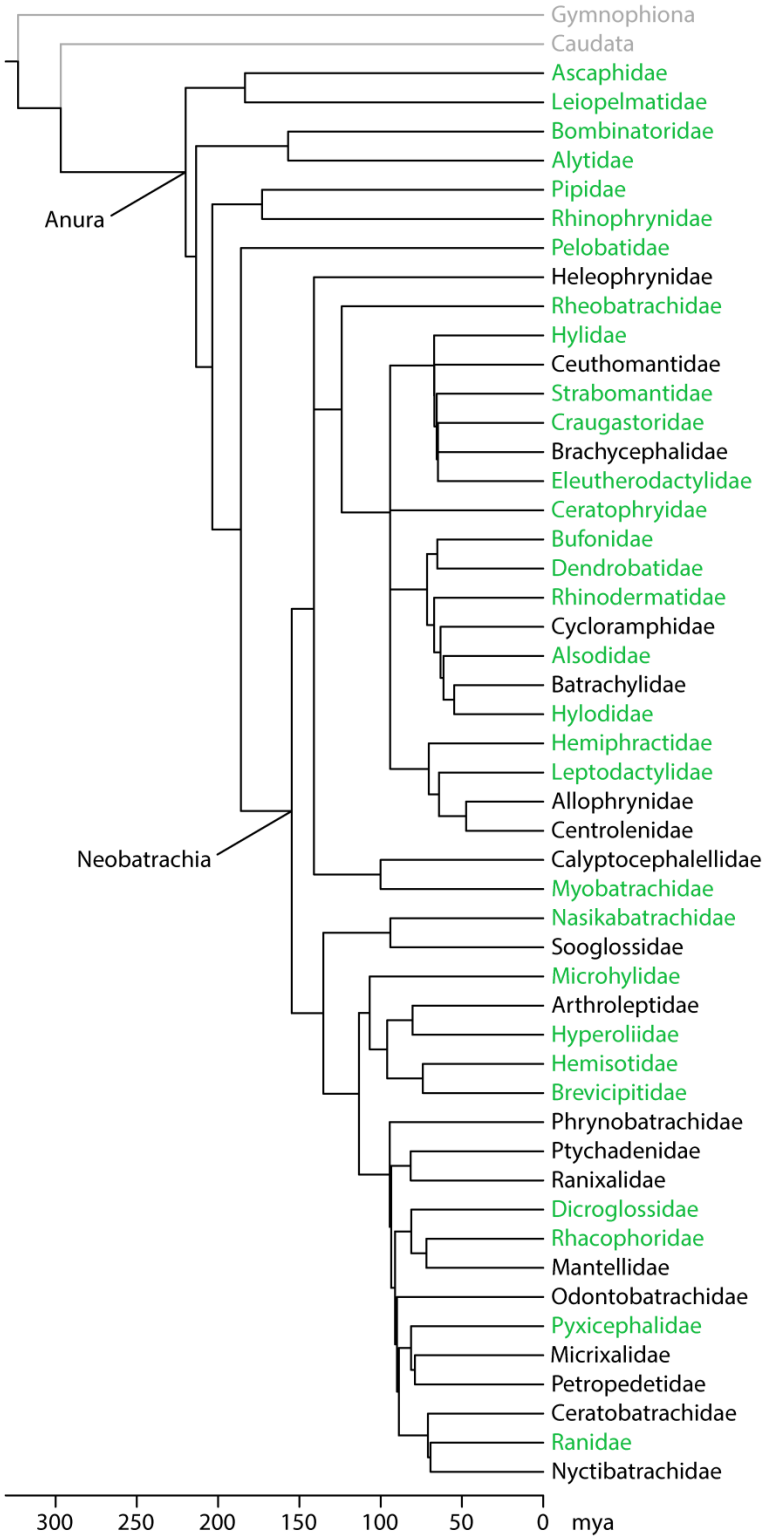


Figure 1 Phylogeny of extant amphibians (extracted from timetree.org, accessed 28.08.2020; Kumar et al. 2017). Anura represented at the family level (black, green), others at the order level (gray). Green: family represented by at least one species in at least one chapter of this thesis; black: family not represented in this thesis.

Some arboreal species are also capable of parachuting or gliding with these two behaviors being defined as, if falling, descending along a path that, respectively, deviates less or more than 45° from the vertical (Oliver, 1951). These various locomotor behaviors allow the anuran

specimens, for example, to access the resources of their respective habitat, to escape from predators, or to encounter mates (Nathan et al. 2008; Liedvogel et al. 2013). Given that the performance capacities of a specimen are determined by its anatomy (Wainwright, 2007), those anatomical features that allow for a high locomotor performance should be selected for which would result in anatomical adaptations to the locomotion.

Origin of jumping and evolution of locomotor modes within the Anura

The evolutionary transition of fins into limbs was accompanied by a shift of the locomotor predominance from the pectoral to the pelvic appendage. In accordance with this, the evolution of the forelimbs initially was ahead of the hindlimbs until the pattern reversed with the occurrence of *Acanthostega* and onwards (Coates et al. 2002; Boisvert, 2005). Fossil evidence indicates that the locomotor behavior of stem-group tetrapods like *Ichthyostega* may have differed from the locomotor modes observed in modern tetrapods and that the early tetrapods rather moved by forelimb "crutching" as, for example, seen in mudskippers (Pierce et al. 2012; Pierce et al. 2013).

Several studies have suggested that a combination of walking and hopping or jumping is the ancestral locomotor behavior of anurans (Přikryl et al. 2009; Reilly & Jorgensen, 2011). This mode of locomotion has evolved from a locomotor pattern that was similar to the one observed in salamanders, that is a lateral undulatory motion combined with alternating limb movements (Lires et al. 2016). Stem-group anurans likely were poor jumpers compared to modern, especially neobatrachian, frogs and toads (Herrel et al. 2016). Jumping, or more precisely the ability for a rapid forward movement by synchronous activation of the hindlimbs, was suggested to have either evolved as an escape mechanism to return from the terrestrial environment into water (Gans & Parsons, 1966; Essner et al. 2010) or as a means to rapidly approach prey (Gans & Parsons, 1966; Reilly et al. 2015). Anyway, it is generally accepted that the jumping ability evolved in riparian species and was accompaigned by skeletal adaptations that resulted in the specific anatomy that characteristic for anurans: The elongated hindlimbs expand the acceleration phase during take-off and therefore maximize the amount of kinetic energy that is transferred to the trunk (e.g., Gans & Parsons, 1966) and the ilio-sacral complex together with the urostyl and the shortened vertebral coulomb allow for an efficient transmission of the forces produced by the hindlimbs (e.g., Gans & Parsons, 1966; Shubin & Jenkins, 1995; Jenkins & Shubin, 1998). In contrast to this, an alternative

hypothesis (Lires et al. 2016) suggests an aquatic origin for the typical anuran Bauplan independent of the evolution of jumping ability.

Leiopelmatid and ascaphid frogs, which together form the sister group of all other anurans (Fig. 1), show no attempt to use the forelimbs to decelerate the body during landing after a jump, but instead hit the ground with a "belly-flop" (Essner et al. 2010). A coordinated landing that involves the forelimbs to absorb and transmit the landing forces has therefore been suggested to have evolved after jumping ability (Essner et al. 2010). An alternative hypothesis supported by fossil evidence, in contrast, assumed the "belly-flop" landing of the Ascaphidae and Leiopelmatidae to be derived and not ancestral (Sigurdsen et al. 2012).

Most anurans swim by synchronous hindlimb movements, whereas the hindlimbs of leiopelmatid and ascaphid species move asynchronously during aquatic locomotion (Abourachid & Green, 1999). The phylogenetic distribution of the motion patterns employed for swimming has led to the assumption that swimming behavior evolved independent from the jumping ability (Abourachid & Green, 1999), although both locomotor modes are similar with regard to hindlimb kinematics (Peters et al. 1996) and activity pattern in the ilio-sacral musculature (Emerson & de Jongh, 1980).

Backward and headfirst burrowing each evolved several times independently within the Anura with headfirst burrowing in most cases being preceded by backward burrowing in an ancestor (Nomura et al. 2009). The anatomical peculiarities that likely are adaptive for the jumping locomotion of anurans were suggested to be, with minor modifications, well-suited and therefore pre-adaptive for backward burrowing (Emerson, 1976). Backward burrowing, as well as the use of quadrupedal walking, might have evolved to explore underground food resources like ant or termits' nests (Emerson, 1976). But backward burrowing also permits the specimens to avoid heat, aridity, predators, or cold winters (Menzies & Tyler, 1977; Wells, 2007). Headfirst burrowing is specie-specifically performed by moving the soil either with the forelimbs or the head (Emerson, 1976; Nomura et al. 2009). It might have evolved to maximize the foraging and prey capturing success underground (Brown et al. 1972; Brown, 1978). With regard to the genus *Alytes*, in which the males carry the eggs wrapped around the hindlimbs, headfirst burrowing likely evolved to prevent the eggs from being damaged during digging (Brown & Crespo, 2000).

Similar to the burrowing locomotion, climbing evolved several times independently, but only within the Neobatrachia (Reilly & Jorgensen, 2011). Some of the arboreal species additionally show parachuting or gliding behavior (Oliver, 1951), which possibly evolved as a means of rapid descent to reach diurnal retreats after foraging in the canopy (Stewart, 1985),

or, in explosively breeding species, to synchronously arrive at the breeding sites near the ground (Wells, 2007).

Anatomical correlates and adaptations to locomotor behaviors

The evolution of new locomotor behavior within the Anura likely was accompanied by anatomical adaptations to satisfy the biomechanical demands presumably imposed by the new motion patterns. And in fact, correlations of anatomical traits with locomotor behavior or performance have been reported in a huge number of studies. Various studies, for example, found that jumping, walking, swimming, and burrowing behavior were each associated with specific proportions of the hindlimbs relative to the forelimbs or snout-vent-length, or of the long bones within limbs (e.g., Laurent, 1964; Brown et al. 1972; Zug, 1972; Emerson, 1978; Enriquez-Urzelai et al. 2015; Astley, 2016; Gómez & Lires, 2019; Moen, 2019). Likewise, the anatomy of the ilio-sacral-complex (including the urostyl) differed between species of certain locomotor groups (e.g., Emerson, 1979; Reilly & Jorgensen, 2011; Jorgensen & Reilly, 2013). A higher jumping performance was associated with larger hindlimb muscle and specific physiological muscle properties (Chadwell et al. 2002; Astley, 2016). Frequently swimming species had hindlimbs with a higher relative muscle mass (Moen, 2019).

Extensive foot webbing was suggested to be adaptive for swimming and might also co-occur in concert with hand webbing in climbing species (Laurent, 1964). Adaptations to climbing include the presence of enlarged finger and toe tips (Moen et al. 2013) with adhesive pads (Noble & Jaeckle, 1928; Emerson & Diehl, 1980) and modifications for the distal forelimbs for grasping (Manzano et al. 2008). The presence of well-developed metatarsal tubercles and relatively short hindlimbs (Emerson, 1976), robust prehalluxes (Kley & Kearney, 2006), as well as modifications of feed muscles (Sanders & Davies, 1983; Burton, 2001; Blotto et al. 2017) have been interpreted as an adaptation to backward burrowing. In contrast, headfirst burrowing was associated with modified skulls (Menzies & Tyler, 1977; Davies, 1984), massive mandibles (Menzies & Tyler, 1977), a specific humerus shape comprising a large crista ventralis (Keeffe & Blackburn, 2020), and modifications of the manus (Kley & Kearney, 2006).

It is striking that most of the studies linking anatomical features with locomotor behavior focused on the hindlimbs or pelvic girdle. The forelimbs, although involved in locomotion (e.g., Sanders & Davies, 1983; Nauwelaerts & Aerts, 2006; Manzano et al. 2008; Reynaga et al. 2018), received less attention. The pectoral girdle that connects the forelimbs

and the axial skeleton was rarely considered with regard to anatomical adaptations to locomotor behavior. A specialized pectoral girdle morphology was thought to be adaptive for headfirst burrowing (Emerson, 1976). One study (Zug, 1972) found that higher jumping capacities were associated with shorter scapula, whereas another (Soliz et al. 2017) reported the opposite. The latter study also observed that the scapulae of species with higher jumping abilities were equipped with broad proximal and distal ends and that the claviculae and coracoids were relatively long compared to the corresponding bones in weaker jumpers. These few and partly contradictory previous reports reveal the need for further studies on the anatomical adaptations of the anuran pectoral girdle to the different locomotor behaviors.

Anatomy of the anuran pectoral girdle

The anatomy of the pectoral girdle skeleton is relatively well studied across anuran taxa (e.g., Procter, 1921; Parker, 1934; Griffiths, 1956/57; Trueb, 1973; Kaplan, 2000). These studies generally described the anuran pectoral girdle as follows (Fig. 2): The pectoral girdle consists of two c-shaped halves that are ventrally overlapping or fused such that the entire girdle is u-shaped. The dorsal part of the glenoid fossa is formed by the bony scapula. The cartilaginous and sometimes partly calcified plate-like suprascapula is located dorsally to the scapula. The anterior margin and the medial and lateral surface of the suprascapula are partly covert by a dermal bone, the cleithrum. Ventral to the scapula, there are two bones roughly running from medial to lateral: the clavicula (dermal bone, anteriorly) and the coracoid (posteriorly). Medially these two bones are connected by a cartilaginous and sometimes partly calcified plate. The anterior part of this cartilage is formed by the procoracoid cartilage, the posterior by the epicoracoid cartilage. The pro- and epicoracoid cartilages of both body halves are species-specifically freely overlapping (aciferal girdle type) or fused to a lesser or greater extend (firmisternal girdle type or intermediate morphologies). There species-specifically also is an episternum (or omosternum) anterior to or/and a sternum posterior to the pro- and epicoracoid cartilages.

Despite the generally conserved morphology of the anuran pectoral girdle skeleton, there are considerable interspecific differences. For example, the ventral notch that separates the pars acromialis from the pars glenoidalis of the scapula in most anurans is secondarily reduced in *Ascaphus truei* (Borsuk-Bialynicka & Evans, 2002). In the Microhylidae, there is a tendency towards the reduction or even loss of the clavicles and procoracoid cartilages (Parker, 1934). In addition, there are major interspecific variations in the shape and/or

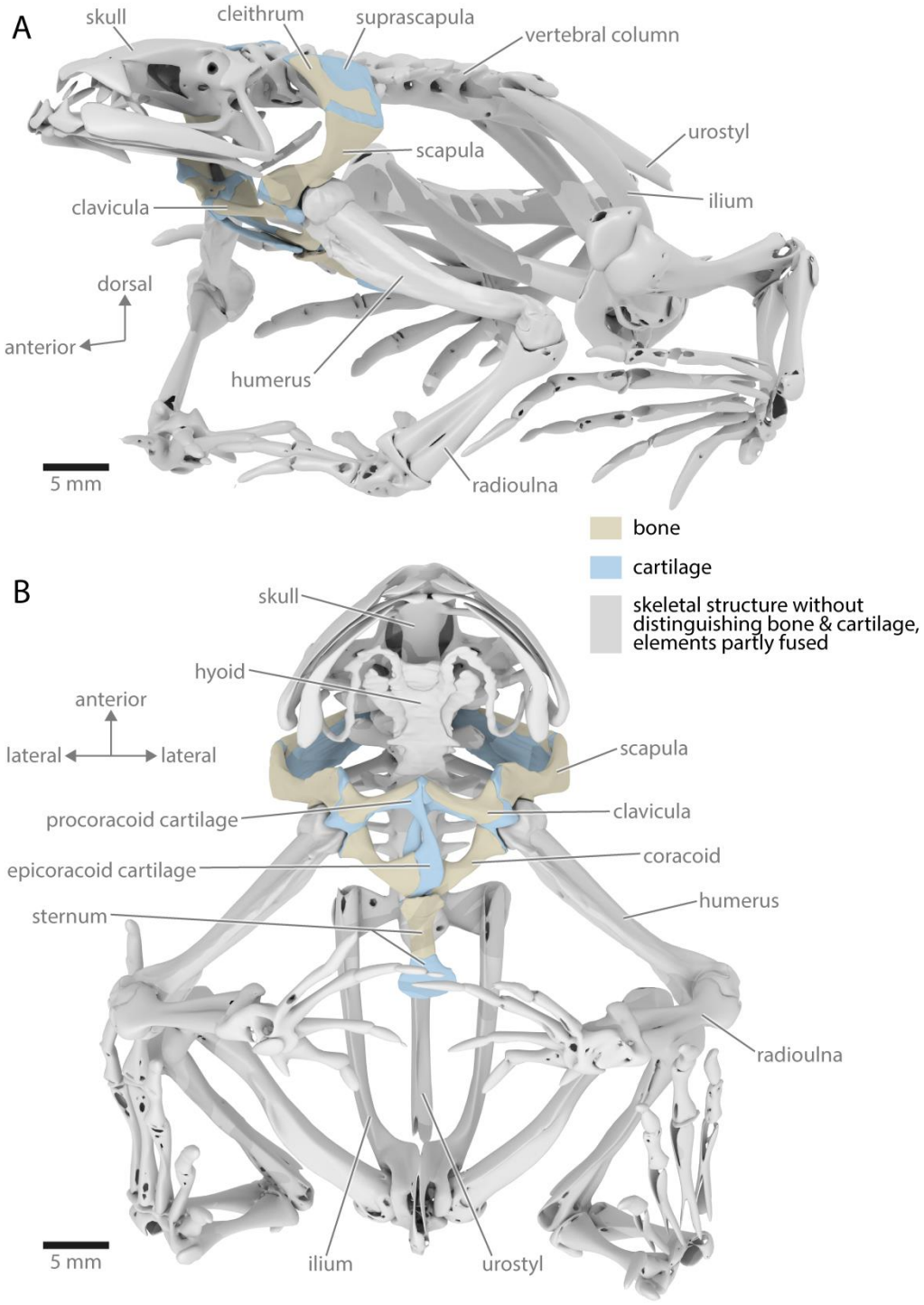


Figure 2 Skeleton of *Bufo bufo* (ZMH A04660) with emphasis on the pectoral girdle. Surface model derived from µCT volume of the unstained specimen. Bone (beige) and cartilage (blue) only separated for elements of the pectoral girdle; other skeletal elements (grey) segmented more coarsely, uncalcified cartilage mostly omitted, elements partly fused, and surfaces containing artifacts like holes. **A** Anterolateral view, anterior approximately to the left. **B** Ventral view, anterior to the top.

orientation of the bones of the pectoral girdle and in the shape of sternum and episternum within the Anura (compare, e.g., Trueb, 1973).

The pectoral girdle muscles, herein defined as all muscle that originate from or insert onto the pectoral girdle (plus the m. latissimus dorsi and the portio abdominalis of the m. pectoralis due to their close association with pectoral girdle muscles), have been described for selected species or groups (e.g., Dugès, 1835; Gaupp, 1896; Bigalke, 1927; Ritland, 1955; Burton, 1983; Manzano, 2000). Several muscles connect the pectoral girdle skeleton with the axial skeleton (skull and vertebral column) and with the forelimb (Fig. 3). Additional pectoral girdle muscles include muscles inserting onto the hyoid apparatus, trunk muscle with attachments to the pectoral girdle skeleton, and there is at least one girdle intern muscle (origin and insertion located on pectoral girdle).

Most of the pectoral girdle muscles or muscle groups have consistently been recognized and named in previous studies but there are inconsistencies with regard to some muscles crossing the shoulder joint: The m. scapulohumeralis profundus anterior, a muscle originating from the anterolateral surface of the scapula and inserting proximal on the humerus, was observed in various anuran species, including representatives of the Ranidae (Tyson, 1987). In a different study (Gaupp, 1896), however, no such muscle was reported in species of the genus *Rana*. Ritland (1955) implied the homology of his m. supracoracoideus superficialis with the m. pectoralis portio epicoracoidea described by Gaupp (1896) and thereby implied the homology of these muscles in *Bufo bufo* und *Rana*. Diogo & Ziermann (2014), in contrast, reported the presence of both, a pars epicoracoidea of the m. pectoralis and a separate m. supracoracoideus, which, in turn, contradicts the homology assumption of Bigalke (1927). These inconsistencies in previous descriptions of the anuran shoulder joint muscles call for further comparative studies to fully understand this anatomical complex. This would not only resolve the inconsistencies in literature accounts, but would also constitute the base for future analyses assessing potential correlations of muscle character states with locomotor behaviors, the reconstruction of muscle evolution within the Anura, and the incorporation of the Anura in current attempts (e.g., Diogo et al. 2016; Molnar et al. 2018) to reconstruct the evolutionary transitions of fins into limbs and of limbs within tetrapods.

Morphological methods including geometric morphometrics

Commonly utilized techniques for anatomical and morphological studies comprise, among others, manual dissection (e.g., Diogo & Ziermann, 2014), histological serial sectioning (e.g., Soliz et al. 2018), episcopic techniques (e.g., Hegre & Brashear, 1946, 1947), and digital dissection of virtual volume representations of specimens (e.g., Cox & Faulkes, 2014;

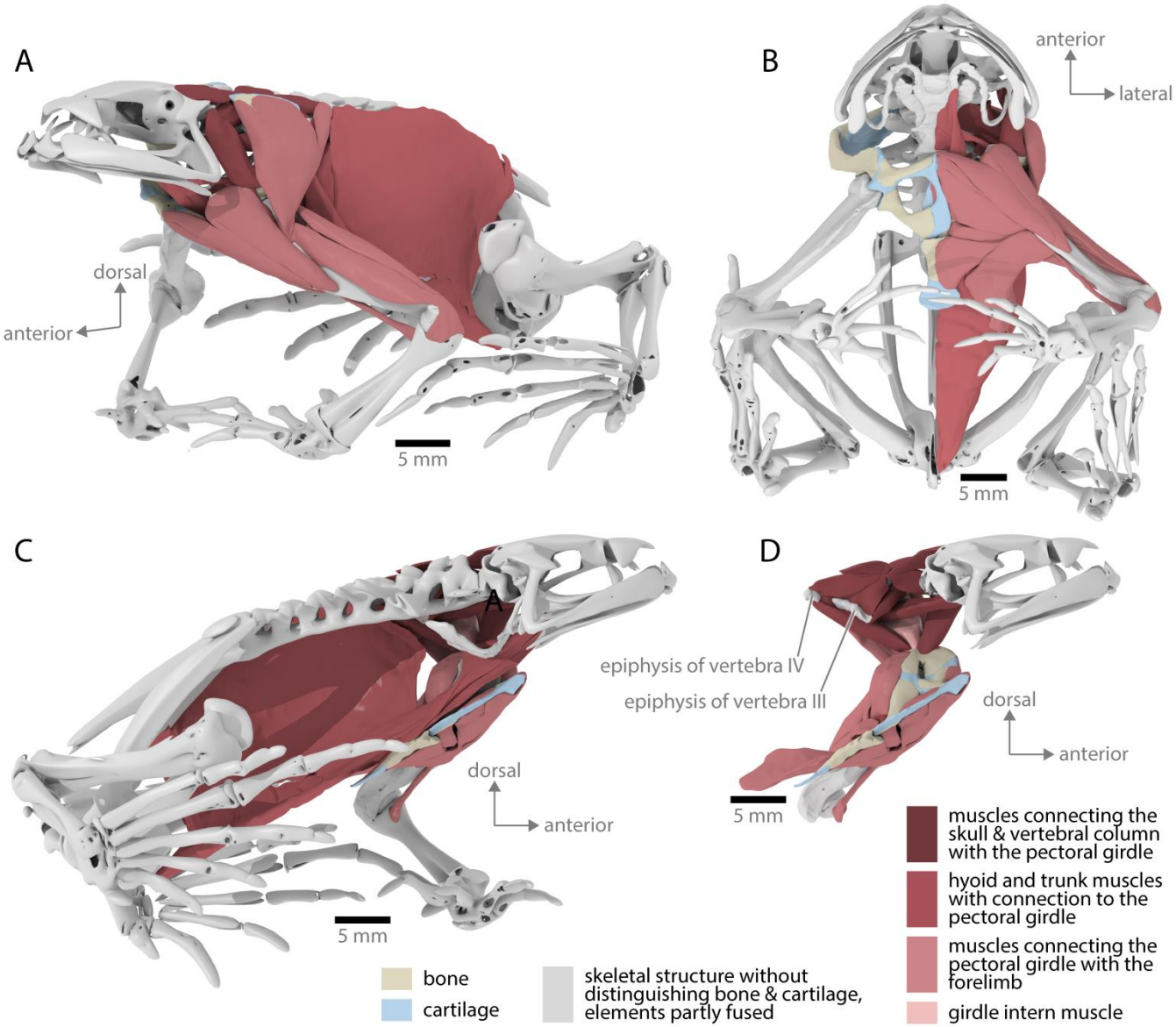


Figure 3 Left-side pectoral girdle muscles of *Bufo bufo* (ZMH A04660). Surfaces of skeletal elements as in Figure 2; surfaces of muscles (red) derived from iodine-contrast-enhanced µCT volume. **A** Anterolateral view, anterior approximately to the left. **B** Ventral view, anterior to the top. **C** Medial view, anterior to the right. Right pectoral girdle half and right forelimb removed. **D** Same as C, but left radioulna, left hand, vertebral column, pelvic girdle, hindlimbs, hyoid muscles, and trunk muscles removed.

Lautenschlager et al. 2014). With regard to the latter, the volumes for digital dissection can be acquired by (micro-)computed tomography ([µ]CT) (e.g., Quayle et al. 2014; Heiss et al. 2016), magnetic resonance imaging (MRI; e.g., Sharp & Trusler, 2015; Klinkhamer et al. 2017), digitally stacking histological serial sections (e.g., Pomikal et al. 2011; Henne et al. 2017), or episcopic imaging (e.g., Weninger et al. 1998, 2006). Each of the approaches for volume data generation, as well as each of the methods for anatomical studies, comes with its own advantages, disadvantages, and limitations. Most methods, for example, involve the destruction of the specimen, which might be problematic for rare museum specimens.

Histological serial sectioning and episcopic imaging are limited by the specimen size as the embedded specimen has to fit within the microtome. Among the approaches of volume data generation for digital dissection, virtually stacked histological serial sections usually allow for the highest resolutions (at least in the section plane) and tissue contrasts, but the volumes might be affected by distortions and alignment artifacts (e.g., Streicher et al. 1997; Malandain et al. 2004). MicroCT scanning and MRI offer the advantages that the specimens are not destructed during volume data acquisition, that data acquisition is comparably easy and fast, that the volumes are inherently aligned, that anatomical structures are depicted in their natural position and shape (as long as no artefactual deformations do occur during specimen preparation, compare, e.g., Buytaert et al. 2014; Hedrick et al. 2018), and that isotropic voxel sizes might be obtained. The combination of these advantages and the increasing accessibility of µCT scanners likely is the reason why volumes acquired by µCT presumably are the most widely used data type for digital dissection (e.g., Cox & Faulkes, 2014; Quayle et al. 2014; also see Koç et al. 2019).

Calcified tissues can directly be visualized in CT volumes while the visualization of soft tissues generally requires some kind of contrast enhancement; iodine-based contrast staining seems to be the most common approach for soft tissue contrast enhancement in vertebrates (compare, e.g., Gignac et al. 2016; Li et al. 2016). But even with iodine-staining some tissues like cartilage, tendons, blood vessels, and minor nerves often remain barely visible or invisible in CT volumes (personal observation; Lautenschlager et al. 2014; Bribiesca-Contreras & Sellers, 2017; Sullivan et al. 2019). Alternative approaches of volume data acquisition would need to be considered if organs or elements constituted by any of these tissue types were of importance for the anatomical work.

A digital dissection usually involves the identification and separation (segmentation) of the anatomical structures in the volume data (compare, e.g., Cox & Faulkes, 2014; Lautenschlager et al. 2014). The results of such a process often are used to generate polymesh surfaces, which, in turn, may serve to illustrate anatomical findings in publications and may be distributed as three-dimensional (3d) interactive portable document format (PDF) files (e.g., Ruthensteiner & Heß, 2008; Lautenschlager, 2014) or as web-based figures (Quayle et al. 2014). Furthermore, surfaces derived from digital dissections can be used in advanced analyses like geometric morphometrics (e.g., Kesterke et al. 2018), multibody dynamic analyses (MDA; e.g., Curtis et al. 2010), finite element analyses (FEA; e.g., Fortuny et al. 2015), and computational fluid dynamics (CDF; e.g., Hammel et al. 2013).

Geometric morphometrics refers to the landmark-based analysis of shape with shape being defined as all geometric information that remains after the scale, the position, and the orientation in space have been removed (Zelditch et al. 2012). The landmarks are represented by two- or three-dimensional Cartesian coordinates that, like almost all kinds of measurements, are affected by measurement error (Arnqvist & Mårtensson, 1998; Fruciano, 2016). Considerable or significant artefactual variance in landmark data has been shown to be caused by inter- and intra-observer variation (e.g., Ross & Williams, 2008; Barbeito-Andrés et al. 2012; Robinson & Terhune, 2017), the reduction of an 3d specimen to a 2d representation (e.g., Cardini, 2014; Buser et al. 2018), and the choice of the approach for landmark data acquisition (e.g., Hale et al. 2014; Fruciano et al. 2017; Robinson & Terhune, 2017; Shearer et al. 2017; Marcy et al. 2018). In addition, some sources of measurement error specific to CT-based geometric morphometric have previously been identified: The choice of the filter during CT scanning (Simon & Marroig, 2015), the spatial resolution of the CT volume (Gunz et al. 2012), and the threshold selected for segmentation (Williams & Richtsmeier, 2003) potentially introduce considerable artefactual variance in the landmark data. These previous studies highlight the importance of assessing and controlling the measurement error in geometric morphometric studies (also see Klingenberg, 2015; Fruciano, 2016; Robinson & Terhune, 2017).

Once landmark data are acquired, shape differences can be visualized and a huge variety of analyses can be performed (e.g., Zelditch et al. 2012; Slice, 2005). The data can, for example, be used in multivariate statistical tests for shape differences between groups in a phylogenetic context (e.g., Adams & Collyer, 2018a,b) or in test for allometric effects (e.g., Klingenberg, 2016). Additional analyses include, but are not limited to, tests for modularity (i.e., the presence of highly correlated subsets of landmarks coordinates, whereas the covariation between such subsets is relatively weak; Schlosser, 2002; Adams, 2016) or the analysis of asymmetry (e.g., Klingenberg, 2015).

Aims of this study

Previous studies have shown that various anatomical traits of anurans are associated with or adaptive for specific locomotor modes. This is likely also true for the pectoral girdle, but has rarely been investigated. Therefore, the first aim of this study is to resolve the relationships between locomotor behavior, anatomy, and biomechanical function of the anuran pectoral girdle for various species distribute across the anuran phylogeny (Fig. 1). Given that motion

arises from the interaction of skeletal elements and muscles and assuming that the skeletal geometry mostly determines the effects and effectiveness of muscles, assessing the relation of pectoral girdle shapes with the locomotor modes by means of geometric morphometrics and performing subsequent biomechanical simulations seems to be a practicable approach to achieving the first aim. In a first step, the applicability of µCT-based 3d geometric morphometrics for the analysis of the pectoral girdle shape will be tested by determining the measurement error that is introduced in the steps of surface generation and landmark acquisition (Chapter two). This analysis will show that a significant amount of artefactual variance is introduced in the steps of segmentation and surface generation. The accuracy of different CT segmentation and surface generation approaches will therefore be assess in order to identify the approach that results in the most accurate surfaces (Chapter three). The insights of these previous analyses will then be applied for the analysis of the anuran pectoral girdle shape and its relation to locomotor behavior (Chapter four).

The inconsistencies in the identification of and terminology for the anuran shoulder joint muscles in the literature demonstrate the need for further comparative studies that identify the distinct muscle units and establish the inter-specific homologies of these muscles, which is the second aim herein. To do so, the workflow of episcopic microscopy (Weninger et al. 1998) will first be adapted for frog-sized specimens (Chapter five). This will allow for the consideration and visualization of structures (i.e., nerves and tendons) that are barely visible in contrast-enhanced CT volumes. Finally, the ontogenetic development of the shoulder joint muscles and their relative position to each other and to nerves will be assessed in order to identify the muscle units crossing the shoulder joint and to establish their interspecific homology including the suggestion of a consistent nomenclature (Chapter six).

References

Abourachid A, Green DM (1999) Origins of the frog-kick? Alternate-leg swimming in primitive frogs, families Leiopelmatidae and Ascaphidae. J Herpetol **33**, 657.

Adams DC (2016) Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. Methods Ecol Evol **7**, 565–572.

Adams DC, Collyer ML (2018a) Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst Biol **67**, 14–31.

Adams DC, Collyer ML (2018b) Phylogenetic ANOVA: group-clade aggregation, biological challenges, and a refined permutation procedure. Evolution; international journal of organic evolution **72**, 1204–1215.

Ahn AN, Monti RJ, Biewener AA (2003) *In vivo* and *in vitro* heterogeneity of segment length changes in the semimembranosus muscle of the toad. J Physiol (Lond) **549**, 877–888.

Arnqvist G, Mårtensson T (1998) Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. Acta Zool Acad Sci Hung **44**, 73–96.

Astley HC (2016) The diversity and evolution of locomotor muscle properties in anurans. J Exp Biol **219**, 3163–3173.

Barbeito-Andrés J, Anzelmo M, Ventrice F, Sardi ML (2012) Measurement error of 3D cranial landmarks of an ontogenetic sample using computed tomography. J Oral Biol Craniofac Res **2**, 77–82.

Bigalke R (1927) Zur Myologie der Erdkröte. (*Bufo vulgaris*, Laurenti.). Z Anat Entwickl Gesch **82**, 286–353.

Blotto BL, Pereyra MO, Faivovich J, Dias PHDS, Grant T (2017) Concentrated evolutionary novelties in the foot musculature of Odontophrynidae (Anura: Neobatrachia), with comments on adaptations for burrowing. Zootaxa **4258**, 425–442.

Boisvert CA (2005) The pelvic fin and girdle of *Panderichthys* and the origin of tetrapod locomotion. Nature **438**, 1145–1147.

Borsuk-Bialynicka M, Evans SE (2002) The scapulocoracoid of an Early Triassic stem-frog from Poland. Acta Palaeontol Pol **47**, 79–96.

Bribiesca-Contreras F, Sellers WI (2017) Three-dimensional visualisation of the internal anatomy of the sparrowhawk (*Accipiter nisus*) forelimb using contrast-enhanced micro-computed tomography. PeerJ **5**, e3039.

Brown LE (1978) Subterranean feeding by the chorus frog *Pseudacris streckeri* (Anura: Hylidae). Herpetologica **34**, 212–216.

Brown LE, Crespo EG (2000) Burrowing behavior of the midwife toads *Alytes cisternasii* and *Alytes obstetricans* (Anura, Discoglossidae). Alytes **17**, 101–113.

Brown LE, Jackson HO, Brown JR (1972) Burrowing behavior of the chorus frog, *Pseudacris streckeri*. Herpetologica **28**, 325–328.

Burton TC (1983) The musculature of the papuan frog *Phrynomantis stictogaster* (Anura, Microhylidae). J Morphol **175**, 307–324.

Burton TC (2001) Variation in the foot muscles of frogs of the family Myobatrachidae. Aust J Zool **49**, 539.

Buser TJ, Sidlauskas BL, Summers AP (2018) 2D or not 2D? Testing the utility of 2D vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). Anat Rec **301**, 806–818.

Buytaert J, Goyens J, Greef D de, **Aerts P, Dirckx J** (2014) Volume shrinkage of bone, brain and muscle tissue in sample preparation for micro-CT and light sheet fluorescence microscopy (LSFM). Microsc Microanal **20**, 1208–1217.

Cardini A (2014) Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures? Hystrix **25**, 73–81.

Chadwell BA, Hartwell HJ, Peters SE (2002) Comparison of isometric contractile properties in hindlimb extensor muscles of the frogs *Rana pipiens* and *Bufo marinus*: functional correlations with differences in hopping performance. J Morphol **251**, 309–322.

Coates MI, Jeffery JE, Rut M (2002) Fins to limbs: what the fossils say. Evol Dev **4**, 390–401.

Cox PG, Faulkes CG (2014) Digital dissection of the masticatory muscles of the naked mole-rat, *Heterocephalus glaber* (Mammalia, Rodentia). PeerJ **2**, e448.

Curtis N, Jones MEH, Evans SE, Shi J, O'Higgins P, Fagan MJ (2010) Predicting muscle activation patterns from motion and anatomy: modelling the skull of *Sphenodon* (Diapsida: Rhynchocephalia). J R Soc Interface **7**, 153–160.

Davies M (1984) Osteology of the myobatrachine frog *Arenophryne rotunda* Tyler (Anura: Leptodactylidae) and comparisons with other myobatrachine genera. Aust J Zool **32**, 789–802.

Diogo R, Johnston P, Molnar JL, Esteve-Altava B (2016) Characteristic tetrapod musculoskeletal limb phenotype emerged more than 400 MYA in basal lobe-finned fishes. Sci Rep **6**, 37592.

Diogo R, Ziermann JM (2014) Development of fore- and hindlimb muscles in frogs: morphogenesis, homeotic transformations, digit reduction, and the forelimb-hindlimb enigma. J Exp Zool B Mol Dev Evol **322**, 86–105.

Duellman WE, Trueb L (1994) Biology of amphibians. Baltimore: Johns Hopkins University Press.

Dugès A (1835) Recherches sur l'ostéologie et la myologie des Batraciens à leurs différens âges: Présentées à l'Académie Royale des Sciences. Paris: J.-B. Baillière.

Emerson SB (1976) Burrowing in frogs. J Morphol **149**, 437–458.

Emerson SB (1978) Allometry and jumping in frogs: helping the twain to meet. Evolution **32**, 551–564.

Emerson SB (1979) The ilio-sacral articulation in frogs: form and function. Biol J Linnean Soc **11**, 153–168.

Emerson SB, de Jongh HJ (1980) Muscle activity at the ilio-sacral articulation of frogs. J Morphol **166**, 129–144.

Emerson SB, Diehl D (1980) Toe pad morphology and mechanisms of sticking in frogs. Biol J Linnean Soc **13**, 199–216.

Enriquez-Urzelai U, Montori A, Llorente GA, Kaliontzopoulou A (2015) Locomotor mode and the evolution of the hindlimb in Western Mediterranean anurans. Evol Biol **42**, 199–209.

Essner RL, Suffian DJ, Bishop PJ, Reilly SM (2010) Landing in basal frogs: evidence of saltational patterns in the evolution of anuran locomotion. Naturwissenschaften **97**, 935–939.

Fortuny J, Marcé-Nogué J, Heiss E, Sanchez M, Gil L, Galobart À (2015) 3D bite modeling and feeding mechanics of the largest living amphibian, the Chinese Giant salamander *Andrias davidianus* (Amphibia: Urodela). PLoS ONE **10**, e0121885.

Frost DR (2020) Amphibian Species of the World: an online reference. Version 6.1; [cited 2020 Aug 26]. Available from: http://research.amnh.org/herpetology/amphibia/index.html.

Fruciano C (2016) Measurement error in geometric morphometrics. Dev Genes Evol **226**, 139–158.

Fruciano C, Celik MA, Butler K, Dooley T, Weisbecker V, Phillips MJ (2017) Sharing is caring? Measurement error and the issues arising from combining 3D morphometric datasets. Ecol Evol **7**, 7034–7046.

Gans C, Parsons TS (1966) On the origin of the jumping mechanism in frogs. Evolution **20**, 92.

Gaupp EWT (1896) A. Ecker's und R. Wiedersheim's Anatomie des Frosches. Erste Abteilung. Lehre vom Skelett und vom Muskelsystem. Braunschweig: Friedrich Vieweg und Sohn.

Gignac PM, Kley NJ, Clarke JA, et al. (2016) Diffusible iodine-based contrast-enhanced computed tomography (diceCT): an emerging tool for rapid, high-resolution, 3-D imaging of metazoan soft tissues. J Anat **228**, 889–909.

Gómez RO, Lires AI (2019) High ecomorphological diversity among Early Cretaceous frogs from a large subtropical wetland of Iberia. Comptes Rendus Palevol.

Griffiths I (1956/57) The structure and development of the breast-shoulder apparatus in Amhibia, Salientia. PhD Thesis. London.

Gunz P, Ramsier M, Kuhrig M, Hublin J-J, Spoor F (2012) The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. J Anat **220**, 529–543.

Hale AR, Honeycutt KK, Ross AH (2014) A geometric morphometric validation study of computed tomography–extracted craniofacial landmarks. J Craniofac Surg **25**, 231–237.

Hammel JU, Jahn H, Möckel L, et al. (2013) Computational fluid dynamics of sponge aquiferous systems. HASYLAB Annu Rep.

Hedrick BP, Yohe L, Vander Linden A, et al. (2018) Assessing soft-tissue shrinkage estimates in museum specimens imaged with diffusible iodine-based contrast-enhanced computed tomography (diceCT). Microsc Microanal **417**, 1–8.

Hegre ES, Brashear AD (1946) Block-surface staining. Stain Technol **21**, 161–164.

Hegre ES, Brashear AD (1947) The block-surface method of staining as applied to the study of embryology. Anat Rec **97**, 21–28.

Heiss E, Handschuh S, Aerts P, van Wassenbergh S (2016) Musculoskeletal architecture of the prey capture apparatus in salamandrid newts with multiphasic lifestyle: does anatomy change during the seasonal habitat switches? J Anat **228**, 757–770.

Henne S, Friedrich F, Hammel JU, Sombke A, Schmidt-Rhaesa A (2017) Reconstructing the anterior part of the nervous system of *Gordius aquaticus* (Nematomorpha, Cycloneuralia) by a multimethodological approach. J Morphol **278**, 106–118.

Herrel A, Moureaux C, Laurin M, et al. (2016) Frog origins: inferences based on ancestral reconstructions of locomotor performance and anatomy. Foss Imp, 108–116.

Jenkins FA, Shubin NH (1998) *Prosalirus bitis* and the anuran caudopelvic mechanism. J Vertebr Paleontol **18**, 495–510.

Jetz W, Pyron RA (2018) The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat Ecol Evol **2**, 850–858.

Jorgensen ME, Reilly SM (2013) Phylogenetic patterns of skeletal morphometrics and pelvic traits in relation to locomotor mode in frogs. J Evolution Biol **26**, 929–943.

Kaplan M (2000) The pectoral girdles of *Rana rugulosa* (Ranidae) and *Nesomantis thomasseti* (Sooglossidae). Herpetologica **56**, 188–195.

Keeffe R, Blackburn DC (2020) Comparative morphology of the humerus in forward-burrowing frogs. Biol J Linnean Soc **171**, 293.

Kesterke MJ, Judd MA, Mooney MP, et al. (2018) Maternal environment and craniofacial growth: geometric morphometric analysis of mandibular shape changes with in utero thyroxine overexposure in mice. J Anat **233**, 46–54.

Kley NJ, Kearney M (2006) Adaptations for digging and burrowing. In *Fins into Limbs* (ed Hall BK), pp. 284–309: University of Chicago Press.

Klingenberg C (2015) Analyzing fluctuating asymmetry with geometric morphometrics: concepts, methods, and applications. Symmetry **7**, 843–934.

Klingenberg CP (2016) Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol.

Klinkhamer AJ, Wilhite DR, White MA, Wroe S (2017) Digital dissection and three-dimensional interactive models of limb musculature in the Australian estuarine crocodile (*Crocodylus porosus*). PLoS ONE **12**, e0175079.

Koç MM, Aslan N, Kao AP, Barber AH (2019) Evaluation of X-ray tomography contrast agents: a review of production, protocols, and biological applications. Microsc Res Tech **82**, 812–848.

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Molecular biology and evolution **34**, 1812–1819.

Laurent RF (1964) Adaptive modifications in frogs of an isolated highland fauna in Central Africa. Evolution **18**, 458.

Lautenschlager S (2014) Palaeontology in the third dimension: a comprehensive guide for the integration of three-dimensional content in publications. Paläontol Z **88**, 111–121.

Lautenschlager S, Bright JA, Rayfield EJ (2014) Digital dissection – using contrast-enhanced computed tomography scanning to elucidate hard- and soft-tissue anatomy in the Common Buzzard *Buteo buteo*. J Anat **224**, 412–431.

Li Z, Ketcham RA, Yan FEI, Maisano JA, Clarke JA (2016) Comparison and evaluation of the effectiveness of two approaches of diffusible iodine-based contrast-enhanced computed tomography (diceCT) for avian cephalic material. J Exp Zool B Mol Dev Evol **326**, 352–362.

Liedvogel M, Chapman BB, Muheim R, Åkesson S (2013) The behavioural ecology of animal movement: reflections upon potential synergies. Anim Migr **1**.

Lires AI, Soto IM, Gómez RO (2016) Walk before you jump: new insights on early frog locomotion from the oldest known salientian. Paleobiology **42**, 612–623.

Malandain G, Bardinet É, Nelissen K, Vanduffel W (2004) Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. NeuroImage **23**, 111–127.

Manzano A (2000) Miología pectoral de algunos Centrolenidae (Amphibia: Anura). Cuad Herpetol **14**, 27–45.

Manzano A, Abdala V, Herrel A (2008) Morphology and function of the forelimb in arboreal frogs: specializations for grasping ability? J Anat **213**, 296–307.

Marcy AE, Fruciano C, Phillips MJ, Mardon K, Weisbecker V (2018) Low resolution scans can provide a sufficiently accurate, cost- and time-effective alternative to high resolution scans for 3D shape analyses. PeerJ **6**, e5032.

Menzies JI, Tyler MJ (1977) The systematics and adaptations of some Papuan microhylid frogs which live underground. J Zool **183**, 431–464.

Moen DS (2019) What Determines the Distinct Morphology of Species with a Particular Ecology? The Roles of Many-to-One Mapping and Trade-Offs in the Evolution of Frog Ecomorphology and Performance. Amer Nat **194**, E81-E95.

Moen DS, Irschick DJ, Wiens JJ (2013) Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. Proc R Soc Lond B Biol Sci **280**, 20132156.

Molnar JL, Diogo R, Hutchinson JR, Pierce SE (2018) Reconstructing pectoral appendicular muscle anatomy in fossil fish and tetrapods over the fins-to-limbs transition. Biol Rev **93**, 1077–1107.

Nathan R, Getz WM, Revilla E, et al. (2008) A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci U.S.A. **105**, 19052–19059.

Nauwelaerts S, Aerts P (2006) Take-off and landing forces in jumping frogs. J Exp Biol **209**, 66–77.

Noble GK, Jaeckle ME (1928) The digital pads of the tree frogs. A study of the phylogenesis of an adaptive structure. J Morphol **45**, 259–292.

Nomura F, Rossa-Feres DC, Langeani F (2009) Burrowing behavior of *Dermatonotus muelleri* (Anura, Microhylidae) with reference to the origin of the burrowing behavior of Anura. J Ethol **27**, 195–201.

Oliver JA (1951) "Gliding" in amphibians and reptiles, with a remark on an arboreal adaptation in the lizard, *Anolis carolinensis carolinensis* Voigt. Amer Nat **85**, 171–176.

Parker HW (1934) A monograph of the frogs of the family Microhylidae. London: British Museum.

Peters SE, Kamel LT, Bashor DP (1996) Hopping and swimming in the leopard frog, *Rana pipiens*: I. Step cycles and kinematics. J Morphol **230**, 1–16.

Pierce SE, Clack JA, Hutchinson JR (2012) Three-dimensional limb joint mobility in the early tetrapod *Ichthyostega*. Nature **486**, 523–526.

Pierce SE, Hutchinson JR, Clack JA (2013) Historical perspectives on the evolution of tetrapodomorph movement. Integr Comp Biol **53**, 209–223.

Pomikal C, Blumer R, Streicher J (2011) Four-dimensional analysis of early pelvic girdle development in *Rana temporaria*. J Morphol **272**, 287–301.

Přikryl T, Aerts P, Havelková P, Herrel A, Roček Z (2009) Pelvic and thigh musculature in frogs (Anura) and origin of anuran jumping locomotion. J Anat **214**, 100–139.

Procter JB (1921) On the variation of the scapula in the Batrachian groups Aglossa and Arcifera. Proc Zool Soc Lond **91**, 197–214.

Púgener LA, Maglia AM (2009) Developmental Evolution of the Anuran Sacro-Urostylic Complex. South Am J Herpetol **4**, 193–209.

Pyron AR, Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol **61**, 543–583.

Quayle MR, Barnes DG, Kaluza OL, McHenry CR (2014) An interactive three dimensional approach to anatomical description-the jaw musculature of the Australian laughing kookaburra (*Dacelo novaeguineae*). PeerJ **2**, e355.

Reilly S, Essner R, Jr. Wren S, Easton L, Bishop PJ (2015) Movement patterns in leiopelmatid frogs: insights into the locomotor repertoire of basal anurans. Behav Process **121**, 43–53.

Reilly SM, Jorgensen ME (2011) The evolution of jumping in frogs: morphological evidence for the basal anuran locomotor condition and the radiation of locomotor systems in crown group anurans. J Morphol **272**, 149–168.

Reynaga CM, Astley HC, Azizi E (2018) Morphological and kinematic specializations of walking frogs. J Exp Zool A Ecol Integr Physiol **329**, 87–98.

Ritland RM (1955) Studies on the post-cranial morphology of *Ascaphus truei*. II. Myology. J Morphol **97**, 215–282.

Robinson C, Terhune CE (2017) Error in geometric morphometric data collection: combining data from multiple sources. Am J Phys Anthropol **164**, 62–75.

Ross AH, Williams S (2008) Testing repeatability and error of coordinate landmark data acquired from crania. J Forensic Sci **53**, 782–785.

Ruthensteiner B, Heß M (2008) Embedding 3D models of biological specimens in PDF publications. Microsc Res Tech **71**, 778–786.

Sanders J, Davies M (1983) Burrowing behaviour and associated hindlimb myology in some Australian hylid and leptodactylid frogs. Aust Zool **21**, 123–142.

Schlosser G (2002) Modularity and the units of evolution. Theory Biosci **121**, 1–80.

Sharp AC, Trusler PW (2015) Morphology of the jaw-closing musculature in the common wombat (*Vombatus ursinus*) using digital dissection and magnetic resonance imaging. PLoS ONE **10**, e0117730.

Shearer BM, Cooke SB, Halenar LB, et al. (2017) Evaluating causes of error in landmark-based data collection using scanners. PLoS ONE **12**, e0187452.

Shubin NH, Jenkins FA (1995) An Early Jurassic jumping frog. Nature **377**, 49–52.

Sigurdsen T, Green DM, Bishop PJ (2012) Did *Triadobatrachus* jump? Morphology and evolution of the anuran forelimb in relation to locomotion in early Salientians. Fieldiana Life Earth Sci **5**, 77–89.

Simon MN, Marroig G (2015) Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomography and the open-source TINA Manual Landmarking Tool software. Front Zool **12**, 12.

Slice DE (2005) Modern morphometrics. In *Modern morphometrics in physical anthropology* (ed Slice DE), pp. 1–45. New York: Kluwer Academic/Plenum Publishers.

Soliz M, Tulli MJ, Abdala V (2017) Osteological postcranial traits in hylid anurans indicate a morphological continuum between swimming and jumping locomotor modes. J Morphol.

Soliz MC, Ponssa ML, Abdala V (2018) Comparative anatomy and development of pectoral and pelvic girdles in hylid anurans. J Morphol **217**, 536.

Stewart MM (1985) Arboreal habitat use and parachuting by a subtropical forest frog. J Herpetol **19**, 391–401.

Streicher J, Weninger WJ, Müller GB (1997) External marker-based automatic congruencing: a new method of 3D reconstruction from serial sections. Anat Rec **248**, 583–602.

Sullivan SP, McGechie FR, Middleton KM, Holliday CM (2019) 3D muscle architecture of the pectoral muscles of European Starling (*Sturnus vulgaris*). Integr Org Biol **1**, oby010.

Trueb L (1973) Bones, Frogs, and evolution. In *Evolutionary biology of the anurans. Contemporary research on major problems* (ed Vial JL), pp. 65–132. Columbia: University of Missouri Press.

Tyson H (1987) The structure and development of the anuran breast-shoulder apparatus, forelimb, and associated musculature. PhD Thesis. Edmonton.

Wainwright PC (2007) Functional versus morphological diversity in macroevolution. Annu Rev Ecol Evol Syst **38**, 381–401.

Wells KD (2007) The ecology & behavior of amphibians. Chicago: University of Chicago Press.

Weninger WJ, Geyer SH, Mohun TJ, et al. (2006) High-resolution episcopic microscopy: a rapid technique for high detailed 3D analysis of gene activity in the context of tissue architecture and morphology. Anat Embryol **211**, 213–221.

Weninger WJ, Meng S, Streicher J, Müller GB (1998) A new episcopic method for rapid 3-D reconstruction: applications in anatomy and embryology. Anat Embryol **197**, 341–348.

Williams FL'E, Richtsmeier JT (2003) Comparison of mandibular landmarks from computed tomography and 3D digitizer data. Clin Anat **16**, 494–500.

Zelditch M, Swiderski DL, Sheets HD (2012) Geometric morphometrics for biologists. a primer. Amsterdam, Boston: Elsevier Academic Press.

Zug GR (1972) Anuran locomotion: structure and function. I. preliminary observations on relation between jumping and osteometrics of appendicular and postaxial skeleton. Copeia **1972**, 613–624.

- Chapter two -

Measurement error in µCT-based three-dimensional geometric morphometrics introduced by surface generation and landmark data acquisition

Karolin Engelkes[1], Jennice Helfsgott[1], Jörg U. Hammel[2,3], Sebastian Büsse[4], Thomas Kleinteich[5], André Beerlink[6], Stanislav N. Gorb[4], Alexander Haas[1]

1) Center of Natural History (CeNak), Universität Hamburg, Hamburg, Germany
2) Institute of Materials Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany
3) Institut für Zoologie und Evolutionsforschung mit Phyletischem Museum, Ernst-Hackel-Haus und Biologiedidaktik, Friedrich-Schiller-Universität Jena, Jena, Germany
4) Functional Morphology and Biomechanics, Institute of Zoology, Kiel University, Kiel, Germany
5) TPW Prüfzentrum GmbH, Neuss, Germany
6) YXLON International GmbH, Hamburg, Germany

Published in Journal of Anatomy (2019),
volume 235, issue 2, pages 357–378, doi 10.1111/joa.12999.

Contributions of KE:

Study design: ~ 95%

Data collection: CT scanning ~ 5%, segmentation & surface generation 100%, landmark data acquisition ~ 55%

Data analysis: 100%

Manuscript drafting: 100%

Contributions confirmed by supervisor: 

Prof. Dr. Alexander Haas

Abstract

Computed-tomography-derived (CT-derived) polymesh surfaces are widely used in geometric morphometric studies. This approach is inevitably associated with decisions on scanning parameters, resolution, and segmentation strategies. Although the underlying processing steps have been shown to potentially contribute artefactual variance to three-dimensional landmark coordinates, their effects on measurement error have rarely been assessed systematically in CT-based geometric morphometric studies. The present study systematically assessed artefactual variance in landmark data introduced by the use of different voxel sizes, segmentation strategies, surface simplification degrees, and by inter- and intra-observer differences, and compared their magnitude to true biological variation. Multiple CT-derived surface variants of the anuran (Amphibia: Anura) pectoral girdle were generated by systematic changes in the factors that potentially influence the surface geometries. Twenty-four landmarks were repeatedly acquired by different observers. The contribution of all factors to the total variance in the landmark data was assessed using random-factor nested PERMANOVAs. Selected sets of Euclidean distances between landmark sets served further to compare the variance among factor levels. Landmark precision was assessed by landmark standard deviation and compared among observers and days. Results showed that all factors, except for voxel size, significantly contributed to measurement error in at least some of the analyses performed. In total, 6.75% of the variance in landmark data that mimicked a realistic biological study was caused by measurement error. In this landmark dataset, intra-observer error was the major source of artefactual variance followed by inter-observer error; the factor segmentation contributed < 1% and slight surface simplification had no significant effect. Inter-observer error clearly exceeded intra-observer error in a different landmark dataset acquired by six partly inexperienced observers. The results suggest that intra-observer error can potentially be reduced by including a training period prior to the actual landmark acquisition task and by acquiring landmarks in as few sessions as possible. Additionally, the application of moderate and careful surface simplification and, potentially, also the use of case-specific optimal combinations of automatic local thresholding algorithms and parameters for segmentation can help reduce intra-observer error. If landmark data are to be acquired by several observers, it is important to ensure that all observers are consistent in landmark identification. Despite the significant amount of artefactual variance, we have shown that landmark data acquired from microCT-derived surfaces are precise enough to study the shape of anuran pectoral girdles. Yet, a systematic assessment of measurement error is advisable for all geometric morphometric studies.

Key words: landmark precision; measurement error; micro computed tomography; surface simplification; thresholding.

Introduction

Shape analysis and shape comparison by means of landmark-based geometrics morphometrics are well-established in biology and related fields, as can be seen by the numerous books (e.g. Bookstein, 2003; Zelditch et al. 2012), review articles (e.g. Adams et al. 2004, 2013; Mitteroecker et al. 2013), and practical applications (e.g. Klingenberg et al. 2002; Cox et al. 2011; Pujol et al. 2014). In general, the geometry of a specimen is represented by a set of

landmarks, also called landmark configuration. Sets of homologous landmarks of different specimens are then superimposed by scaling, translation, and rotation, which allows for the comparison and analysis of shapes and shape differences among specimens.

Landmarks are represented by either two- or three-dimensional (2D, 3D) Cartesian coordinates depending on the research question and study design. Different ways of landmark acquisition have been documented in literature: 2D coordinates have been acquired from, for example, digitized or digital photographs of the specimens (e.g. Monteiro, 2000; Verhaegen et al. 2007), while 3D coordinates have been measured either directly from the specimens or from digital 3D representations of them (e.g. Collard & O'Higgins, 2001; Heuzé et al. 2016). As with all measurements, landmark coordinates are affected by measurement error (Arnqvist & Mårtensson, 1998), with measurement error being defined as the deviation of the measured value from the true value (Rabinovich, 2006). The presence of such artefactual variance in landmark data has recently been noted to have been overlooked in many geometric morphometric studies (Fruciano, 2016), and different studies have stressed the importance of measurement error assessment (e.g. Klingenberg, 2015; Fruciano, 2016; Robinson & Terhune, 2017).

Several factors have been shown to cause measurement error in landmark data (see also review in Fruciano, 2016). The preservation and preparation of specimens can induce artefactual variance by altering the natural form of the structures of interest (Lee, 1982 [linear measurements]; Bonneau et al. 2012). The variability within repeated measurements performed by the same observer and the variability between different observers can also contribute significantly to measurement error (Ross & Williams, 2008; Robinson & Terhune, 2017; reports of relatively large observer error without tests for statistical significance: Curth et al. 2017; Fruciano et al. 2017; Daboul et al. 2018). Nevertheless, the magnitude of observer error often has been considered small or negligible compared with true biological variability (Richtsmeier et al. 1995; O'Higgins & Jones, 1998; Lockwood et al. 2002; Pujol et al. 2014; Barbeito-Andrés et al. 2016).

The reduction of an actual 3D specimen to a 2D representation has been shown to cause error in landmark data. Depending on the specimens and research questions, this error has been acceptable and negligible in some cases but in others it has had considerable impact on biological inferences (Cardini, 2014; Buser et al. 2018 and references therein). In 2D landmark data that had been acquired from photographs and with related techniques, variation in the placement of the specimens in front of the camera and optical distortions have been identified to contribute to measurement error (see Arnqvist & Mårtensson, 1998 and

references therein for a more detailed discussion of measurement error in 2D-image-based workflows). A measurement error introduced by the projection of 3D data to 2D can be avoided by recording 3D landmark coordinates directly from the specimen or a digital 3D representation of it. Studies on the effect of the choice of a method (including the choice of a device) for 3D landmark data acquisition have reported contradictory results on the significance of measurement error caused by method choice and on the effect of choice of method on observer error (e.g. Hale et al. 2014; Fruciano et al. 2017; Robinson & Terhune, 2017; Shearer et al. 2017; Marcy et al. 2018). Hale et al. (2014), for example, found significant differences between landmark data directly digitized from the specimens and landmark data derived from computed tomography (CT) scans of the same specimens. At the level of individual specimens, however, they found no significant difference between corresponding landmark sets. Shearer et al. (2017) reported no significant difference when comparing landmark data acquired from surfaces of a CT scan and of three different surface scanners. Robinson & Terhune (2017), in contrast, observed small, yet significant differences between sets of linear distances directly measured from the specimens (caliper measurements and digitization of landmarks) and from digital representations of them (landmarks from surfaces of CT and laser scanner). With regard to observer error, Shearer et al. (2017), for example, found no global dependence of observer error on the respective method chosen; yet, in some parts of their analysis they found significant dependencies. It should be noted, however, that differences between studies could be the result of different study designs and statistical approaches.

Although polymesh surfaces derived from CT scans have been widely used for the acquisition of 3D landmark sets (e.g. Kulemeyer et al. 2009; Bilfeld et al. 2013; Wang et al. 2015; Kesterke et al. 2018), measurement error associated with this particular workflow has, to our knowledge, rarely been considered beyond artefactual intra- and inter-observer variance. Observer-introduced variance has been reported by, for example, Valeri et al. (1998) and Barbeito-Andrés et al. (2012). Gunz et al. (2012) found that the spatial resolution of microCT (µCT) scans of the mammalian bony labyrinth and the specific thresholds selected for surface generation from volumetric data affected landmark measurements. Threshold selection as source of measurement error was also noted by Williams & Richtsmeier (2003). Simon & Marroig (2015) found no considerable difference in landmark precision when different voxel sizes were used; however, they identified the use of different filters during µCT scanning as potential source of measurement error.

Studies outside the field of geometric morphometrics indirectly support the notion that methodological decisions related to CT scanning and data processing could contribute to measurement error in landmark data acquired from CT-derived surfaces. Measurements such as linear distances or volumes have changed considerably with varying automatically or manually selected thresholds for surface generation (e.g. Coleman & Colbert, 2007; Parkinson et al. 2008). Voxel size has had a considerable effect on measurements and the effect of the segmentation method seemed to increase with voxel size (Christiansen, 2016). Scanner type and imaging conditions have affected the geometry of surfaces generated from CT scans, although the effects have been small compared with the variance introduced by manual segmentation (Colman et al. 2017). All the factors influencing the segmentation result and the geometry of a CT-derived surface potentially contribute to measurement error in landmark data obtained from such surfaces.

The overall magnitude of artefactual variance in measurements acquired from CT data might be small. For example, linear measurements obtained from CT scans of human skulls were shown to differ insignificantly from corresponding measurements on the original skulls (Lorkiewicz-Muszyńska et al. 2015; but see Hildebolt et al. 1990 and Richtsmeier et al. 1995 for contradictory reports). Furthermore, the variance of linear distances derived from landmarks acquired from repeated scans of the same specimen has been low (Richtsmeier et al. 1995). Finally, various measures of trabecular bone have not differed significantly from measures obtained from histological sections when using an appropriate threshold (Fajardo et al. 2002).

The frequent use of CT-derived surfaces in geometric morphometric studies, the scarce and unsystematic assessment of measurement error related to this particular workflow, and the partly contradictory reports on this topic in literature call for further analyses. The present study aims to investigate systematically the contribution of different factors in surface generation and landmark acquisition to the total variance in landmark data. Measurement error due to voxel size, segmentation strategy (i.e. the use of different thresholds), surface simplification, and inter- and intra-observer differences was assessed and compared with true biological variance due to inter-specimen and inter-specific differences, as well as the variation between body sides of a given specimen. Specimens from two species of the genus *Bombina* (Amphibia: Anura: Bombinatoridae) were chosen for obtaining landmark data from the bones of their pectoral girdles. The pectoral girdle of these species consists of two roughly C-shaped halves (in anterior view) that are ventrally overlapping. Each half comprises four bones (Fig. 1A and Supporting Information Fig. S1) connected by cartilaginous elements

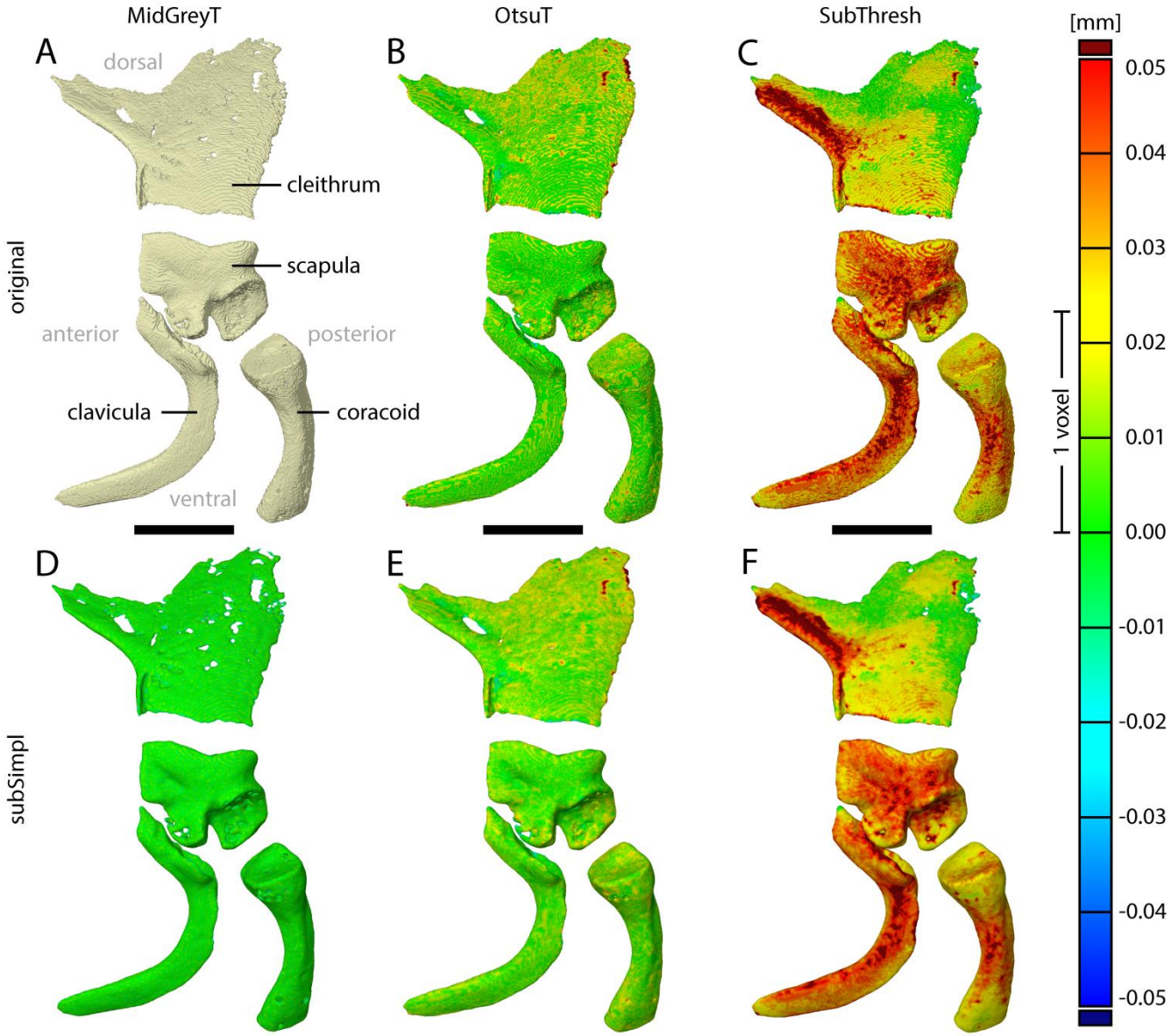


Fig. 1 Selected surface variants of the pectoral girdle bones of *Bombina orientalis* (ZMH A12601, ventral girdle half only), lateral view. All surfaces are based on the same µCT scan. Scale: 2 mm. Color coding denotes distance of the respective surface to the surface in (A). (A) Surface variant generated using the *MidGrey* thresholding algorithm with the intersecting-2-of-3 strategy and no simplification. (B) Surface variant generated using the *Otsu* thresholding algorithm with the intersecting-3 strategy and no simplification. (C) Surface variant generated using a dynamically adapted subjectively optimal threshold and no simplification. (D-F) Subjectively optimal simplified variants of surfaces (A-C), respectively. See text for more details on surface generation.

(Maglia & Púgener, 1998). Recommendations for the reduction of measurement error in landmark data of the anuran pectoral girdle have been derived; these recommendations might be applicable for CT-derived landmark data of other biological structures as well.

Materials and methods

MicroCT scans of several specimens were used to generate different polymesh surface variants of each scan. Those surface variants differed, for example in the underlying segmentation and the degree of surface processing. The surface variants were repeatedly landmarked by different observers to systematically acquire four different sets of landmark configurations ('Landmark Datasets'). Those Landmark Datasets served to analyse different aspects of measurement error.

Specimens and µCT scanning

MicroCT scans of nine specimens of *Bombina orientalis* (Boulenger, 1890) (Amphibia: Anura: Bombinatoridae) and nine of *Bombina bombina* (Linnaeus, 1761) were performed using either a Skyscan 1172 (Bruker microCT), Phoenix Nanotom S (General Electric) or a YXLON FF20 CT or FF35 CT (YXLON International GmbH; Table 1). The specimens were mounted with wadding in plastic containers and CT-scanned in an ethanol-saturated atmosphere. Volumetric datasets were reconstructed from X-ray projections using the reconstruction software delivered with the respective scanner.

The contrast-to-noise ratio (CNR) was calculated for each µCT scan by dividing the difference of the mean gray values of the pectoral girdle bones and the surrounding soft tissues by the standard deviation of the soft tissue grey values. The volume considered for mean bone grey value calculation was determined in Amira® (version 6.0.1; Konrad-Zuse-Zentrum Berlin, FEI Visualization Sciences Group) by manually segmenting the bones (*Magic Wand* tool) and shrinking the selection by two voxels; the volume for calculating the mean value of the soft tissues was arbitrarily chosen. Mean grey values and standard deviations were calculated using the *Material Statistics* module.

The size of the pectoral girdle was recorded for each specimen by determining the distance between the anterodorsal tips of the two scapulae, as well as by averaging the left and right distances between the anteromedial tip of the clavicula and the dorsal end of the respective anterior margin of the cleithrum. Measurements were performed on an *Isosurface* of the result of the MidGreyT segmentation strategy (see below) in Amira®.

Comparison of automatic local thresholding strategies

The anatomical structures of interest in a CT scan need to be segmented before polymesh surfaces of them can be generated. Segmentation can be done by using grey value thresholds

Table 1 Specimens, pectoral girdle sizes, and parameters of µCT scanning.

Species (Collection number)	Girdle size [mm]	Scanner	Current (µA)	Voltage (kV)	Filter	Voxel size [µm]	CNR (bone–soft tissue)
Bombina bombina (ZMH A05110)	11.95 8.09	YXLON FF35 CT	120	100	-	22.75	21.34
Bombina bombina (ZMH A05383)	9.56 6.74	YXLON FF35 CT	120	100	-	22.75	19.93
Bombina bombina (ZMH A05617)	12.02 7.33	YXLON FF20 CT	80	80	-	25.84	48.70
Bombina bombina (ZMH A05619)	10.98 7.70	YXLON FF35 CT	120	100	-	22.75	36.87
Bombina bombina (ZMH A06659)	10.40 8.01	SkyScan1172	100	100	Al 0.5 mm	26.68	16.76
Bombina bombina (ZMH A06683)	10.42 7.55	SkyScan1172	100	100	Al 0.5 mm	26.68	14.81
Bombina bombina (ZMH A06685)	10.57 7.57	SkyScan1172	100	100	Al 0.5 mm	26.68	18.85
Bombina bombina (ZMH A06690)	10.73 8.05	SkyScan1172	100	100	Al 0.5 mm	26.68	20.29
Bombina bombina (ZMH A09674)	7.04 5.71	SkyScan1172	200	49	Al 0.5 mm	21.34	24.98
Bombina orientalis (ZMH A05672)	10.72 7.46	YXLON FF35 CT	120	100	-	30.3	48.88
Bombina orientalis (ZMH A05676)	10.83 7.28	SkyScan1172	100	100	Al 0.5 mm	21.34	11.21
Bombina orientalis (ZMH A05677)	12.00 8.07	YXLON FF35 CT	120	100	-	22.75	33.43
Bombina orientalis (ZMH A05681)	12.97 7.23	YXLON FF35 CT	120	100	-	22.75	33.86
Bombina orientalis (ZMH A05682)	13.60 6.82	YXLON FF35 CT	120	100	-	22.75	35.31
Bombina orientalis (ZMH A12601)	9.53 8.06	Nanotom S	170	60	-	23.37	36.17
Bombina orientalis (ZMH A14347)	12.37 7.18	SkyScan1172	100	100	Al 0.5 mm	26.68	20.85
Bombina orientalis (ZMH A14350)	11.67 6.81	SkyScan1172	100	100	Al 0.5 mm	26.68	20.88
Bombina orientalis (ZMH A14354)	11.97 7.65	SkyScan1172	100	100	Al 0.5 mm	26.68	24.09

CNR, contrast-to-noise ratio calculated by dividing the difference of the mean gray values of pectoral girdle bones and surrounding soft tissues by the standard deviation of soft tissues; Girdle size, first value gives distance between anterodorsal tips of scapulae, second value gives mean distance between anteromedial tip of clavicula and dorsal end of anterior margin of cleithrum.

that allow for the discrimination of different tissue types due to differences in their X-ray absorption. Based on our personal experiences in CT data segmentation, we expect automatic local thresholding algorithms to be superior to automatic global thresholding, and to subjective thresholds determination by eye. However, irrespective of the performance of a given automatic thresholding algorithm, adjacent structures of similar X-ray densities need to be separated manually.

Several automatic local thresholding algorithms are available in the *Auto Local Threshold* plugin (Landini, https://imagej.net/Auto_Local_Threshold, accessed 23 February 2018) for the image processing tool Fiji (based on ImageJ version 1.51n; Schindelin et al. 2012; Schneider et al. 2012). Those algorithms were compared with regard to the quality of the thresholding results in order to determine the best segmentation strategy for the µCT scan of a haphazardly selected *B. orientalis* specimen (ZMH A12601). To do so, a synthetic image stack with defined bone-, soft tissue- and background-areas ('phantom images/stack') was created and virtually CT-scanned (including the simulation of image noise). The thresholding algorithms were applied to the synthetic CT image stack and the thresholding results were then compared with the phantom stack to assess the thresholding quality. Selected algorithm-parameter combinations were applied to resliced versions of the synthetic CT image stack; all thresholding results derived by a given algorithm-parameter combination were combined and the thresholding quality was assessed as above (compare Supporting Information Fig. S2; for a more detailed description of the workflow see Supporting Information Text S1). Assuming that real CT scans behaved similarly to the synthetic CT scan during automatic local thresholding, this approach allowed for the determination of a segmentation strategy that would generate surface geometries close to the real form of the specimens. For the scan of ZMH A12601, the optimal algorithm-parameter combination for the synthetic scan may be expected to result in a surface close to the true shape of the specimen, because the phantom stack has been designed to simulate that particular scan.

The *MidGrey* algorithm with a radius of 9, the parameter -4, and the intersecting-2-of-3 strategy (see Supporting Information Text S1 for explanation) performed best with a misclassification rate of 2.5% in the evaluated volume (Supporting Information Table S1).

Segmentation

All 18 µCT scans were used to compare intra- and inter-specific variation, as well as the variance between both body halves of a given specimen, with the magnitude of measurement error introduced by various factors during surface generation and landmark acquisition (Table 2). To cover a range of surface variants that might reasonably be used in geometric morphometric studies ('reasonable' surfaces), the pectoral girdle bones were segmented (Amira®) three times using different thresholding strategies: the subjectively optimal threshold (*Magic Wand* tool with manual separation of anatomical structures where needed; 'SubThresh') and the result of two different automatic local thresholding algorithms (manual separation of anatomical structures where needed; 'OtsuT', 'MidGreyT'). The subjectively

Table 2 Factors and levels that were considered for evaluating measurement error in relation to inter- and intra-specific variation and the variance between body halves of a given specimen.

Factor	Levels	Abbreviation	Remarks
Species (random)	*Bombina orientalis*	B-ori	
	Bombina bombina	B-bom	
Specimen (random, nested in species)	(9 *B. orientalis* and 9 *B. bombina* specimens)	(species abbreviation and collection-number)	Abbreviated using the abbreviation of the respective species combined with the collection number of the specimen, e.g., B-ori-A12601
Position (random, nested in specimen)	ventral	v	The pectoral girdle half the epicoracoid cartilage of which lays ventral (superficial) to the other half
	dorsal	d	
Segmentation (random, nested in position)	subjectively optimal threshold	SubThresh	
	automatic local thresholding using the *Otsu* algorithm with intersecting-3 strategy	OtsuT	*Otsu* algorithm with *Radius*: 15
	automatic local thresholding using the *MidGrey* algorithm with intersecting-2-of-3 strategy	MidGreyT	*MidGrey* algorithm with *Radius*: 9 and *Parameter 1*: -4
Simplification (random, nested in segmentation)	original, unsimplified surface	original	
	subjectively optimal reduction and smoothing	subSimpl	
Observer (random, nested in simplification)	KE	O1	3 repetitions on three different days (one repetition per day)
	JH	O2	3 repetitions on three different days (one repetition per day)

Factors were considered random and nested in PERMANOVA with the residual term reflecting the repetitions on the same surface variant. Landmark sets acquired from surfaces that were generated according to these factors comprise Landmark Dataset 1.

optimal threshold was determined by eye for each pectoral girdle bone or part of a bone separately. A threshold was considered optimal if the bone-soft-tissue boundary was relatively smooth and laid centric within the grey value gradient between bone and soft tissue voxels. The local version of the algorithm by Otsu (1979) implemented in the *Auto Local Threshold* function was chosen as one of the automatic local thresholding algorithms because local versions of the algorithm by Otsu (1979) yielded good results in previous studies (e.g. Landini et al. 2017; Healy et al. 2018). The three-dimensionality of the data was accounted for by using the intersecting-3 strategy described in Supporting Information Text S1. The *MidGrey* algorithm implemented in the *Auto Local Threshold* function was chosen as a second automatic local thresholding algorithm. It was performed for all specimens with the same parameters and with the strategy that resulted in the best thresholding quality for the reconstructed phantom stack (*Radius*: 9, *Parameter 1*: -4, intersecting-2-of-3 strategy). This segmentation strategy was chosen because it was expected to result in the most natural surfaces for the scan of the specimen selected for phantom stack generation (ZMH A12601) and to produce reasonable surfaces for the other specimens, too.

Table 3 Factors and levels that were considered for evaluating measurement error that was introduced by segmentation and surface generation for *Bombina orientalis* (ZMH A12601; ventral/superficial girdle half only).

Factor	Levels	Abbreviation	Remarks
Downsampling (random)	No downsampling	NoDown	Voxel size: 23.37 µm
	Downsampling of 2x2x2 voxels	Down2	Voxel size: 46.74 µm
	Downsampling of 4x4x4 voxels	Down4	Voxel size: 93.48 µm
Segmentation (random, nested in resolution)	subjectively optimal threshold	SubThresh	
	automatic local thresholding using the *Otsu* algorithm with intersecting-3 strategy	OtsuT	*Otsu* algorithm with *Radius*: 15
	automatic local thresholding using the *MidGrey* algorithm with intersecting-2-of-3 strategy	MidGreyT	*MidGrey* algorithm with *Radius*: 9 and *Parameter 1*: -4
	lowest threshold resulting in a usable surface	minThresh	
	highest threshold resulting in a usable surface	maxThresh	
Simplification (random, nested in segmentation)	original, unsimplified surface	original	
	subjectively optimal reduction and smoothing	subSimpl	
	strong reduction and smoothing	strongSimpl	
Day (random, nested within simplification)	Day 1	Day 1	3 repetitions
	Day 2	Day 2	3 repetitions
	Day 3	Day 3	3 repetitions

Factors were considered random and nested in PERMANOVA with the residual term reflecting the repetitions on the same day. Landmark sets acquired from surfaces that were generated according to these factors comprise Landmark Dataset 2.

The pectoral girdle halves overlap medially in *Bombina*; the girdle half with the ventral (superficial) epicoracoid cartilage of the scan of a selected *B. orientalis* specimen (ZMH A12601) served for testing the effects of more extreme surface variants that would probably not be used in geometric morphometric studies. To simulate different scan resolutions, the voxel size of the µCT scan was decreased (binned) by merging 2x2x2 ('Down2') and 4x4x4 ('Down4') voxels, respectively (*Resample* module in Amira®, filter: *Lanczos*). The bones in the original ('NoDown') and downsampled (Down2, Down4) stacks were segmented five times using the three segmentation strategies described above (SubThresh, MidGreyT, OtsuT), and, additionally, the lowest ('MinThresh') and the highest ('MaxThresh') thresholds that resulted in a usable surface (Table 3).

Surface generation and processing

Different polymesh surface variants were generated from each of the segmentation results for the ventral and dorsal pectoral girdle halves separately and were exported (obj format) in their original condition ('original'). In a next step, copies of these surfaces were simplified (polygon count reduction and smoothing) to a subjective optimal degree that smoothed surface irregularities nicely without losing anatomical details ('subSimpl'). A strongly

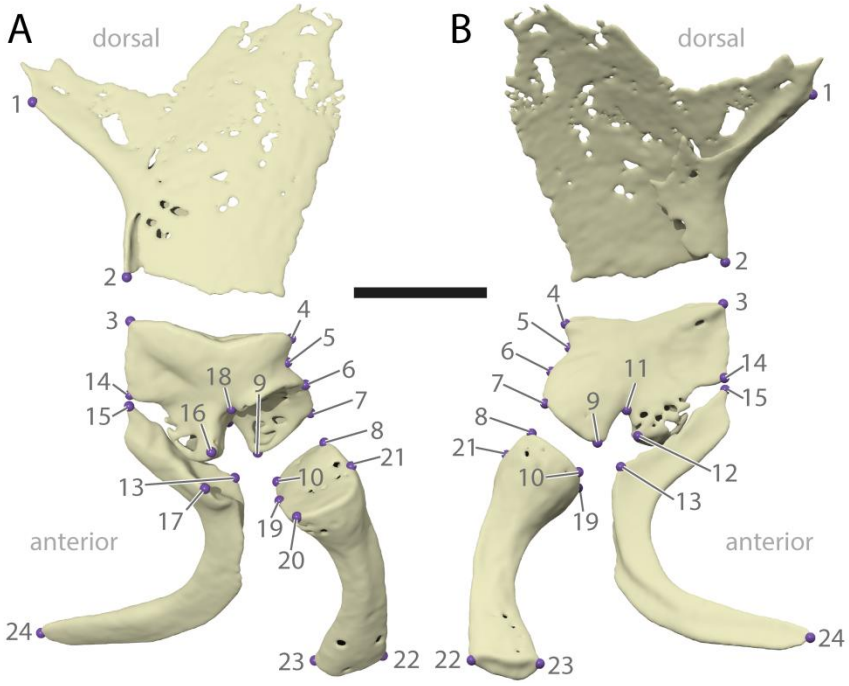


Fig. 2 Landmarks on the girdle half of *Bombina orientalis* (ZMH A12601) that comprises the ventral (superficial) epicoracoid cartilage (MidGreyT, subSimpl surface variant). Scale: 2 mm. (A) Lateral view. (B) Medial view.

simplified surface (highest degree of simplification that resulted in a usable surface; 'strongSimpl') was exported for each segmentation belonging to the ventral girdle half of ZMH A12601 to cover the maximum range of possible surface variants for this specimen. Surface generation, simplification, and export were accelerated using a custom Amira® macro (*MultiExport*, see Engelkes et al. 2018 for details).

The surfaces were converted to ply format in MeshLab (version 1.3.3; Cignoni et al. 2008). Furthermore, the surfaces of the right pectoral girdle halves were mirrored to match the orientation of the left. This allowed for assessing the shape difference of both girdle halves for comparing the magnitude of intra-specimen variation to that of measurement error. The surfaces were mirrored prior to landmark acquisition instead of, as commonly done (seeKlingenberg et al. 2002; Zelditch et al. 2012), mirroring landmark sets. Mirroring surfaces avoided potential bias in landmark acquisition that could have resulted from differing surface orientations.

Landmarks and landmark acquisition

In total, 24 landmarks were defined and acquired in the software Landmark (version 3.0.0.6; Wiley et al. 2005; Fig. 2, Supporting Information Table S2). The landmarks represented the shape of the shoulder girdle bones and their position to each other within the same girdle half. Some of the landmarks (e.g. 5, 11, 20, 21) might as well have been registered as parts of

series of semi-landmarks (Bookstein, 1997; Gunz et al. 2005). Yet, all defined landmarks were landmarks *sensu* Bookstein (2003), and we used them as such to avoid the potential shape variance associated with semi-landmark processing/sliding (compare Perez et al. 2006).

We acquired four different landmark datasets and subjected them to specific analyses to assess different aspects of potential measurement error. Each 'reasonable' surface created according to the factor levels in Table 2 was landmarked by two different observers ('O1', 'O2'), both experienced in landmark acquisition and anuran pectoral girdle anatomy. O1 and O2 initially discussed and agreed on landmark definitions, but then acquired landmark data independently. Each observer landmarked each surface three times with each repetition on a given surface being performed on different days. The total of the landmark sets acquired by O1 and O2 will be called 'Landmark Dataset 1'.

All 45 surfaces of the ventral girdle half of *B. orientalis* specimen ZMH A12601 (Table 3) were landmarked nine times by O1; for each surface variant, the landmark sets were acquired three times repeatedly on three different days ('Landmark Dataset 2').

One of the surfaces of specimen ZMH A12601 (NoDown, MidGreyT, original, ventral girdle half) was landmarked by six different observers (O1, O2, and inexperienced 'O3'–'O6'). Each observer landmarked the surfaces in two sessions on consecutive days with 20 repetitions per session. Inexperienced observers were trained and corrected by O1 during the first five repetitions of the first session. The first five repetitions of each session were discarded, the remaining landmark sets were used for analysis ('Landmark Dataset 3').

The 'reasonable' surface variants (MidGreyT, OtsuT, SubThresh; original, subSimpl) of the ventral girdle halves of two *B. orientalis* specimens were selected based on the CNRs (bone–soft tissue) of the respectively underlying µCT scans. The first specimen (ZMH A05682) was selected because its µCT scan had the CNR (bone–soft tissue) closest to the one of ZMH A12601; the automatic local thresholding algorithms with respective parameters optimal for the scan of ZMH A12601 might thus be expected to be close to optimal for the scan of ZMH A05682, too. The second specimen (ZMH A05676) was chosen because its scan had a CNR (bone–soft tissue) that was the most different from the one of ZMH A12601; thus, the automatic local thresholding algorithms with respective parameters applied might be expected to be suboptimal for the ZMH A05676 scan. The surfaces were landmarked nine times by O1; for each surface variant, the landmark sets were acquired three times repeatedly in three different sessions with at least 10 h between sessions ('Landmark Dataset 4'). Supporting Information Table S3 gives an overview of the composition of Landmark Datasets 1–4.

Superimposition

All surfaces derived from the same µCT scan had the same position in space. Therefore, landmark sets on the surfaces of the same girdle half should, in theory, be in perfect superimposition and all remaining variance in landmark position would have to be the result of measurement error. A superimposition of these landmark sets might mask or alter potential measurement error under this condition (compare Corner et al. 1992; von Cramon-Taubadel et al. 2007; Ross & Williams, 2008). Therefore, we computed the mean landmark configuration of each girdle half in Landmark Dataset 1 and only these mean configurations were superimposed using a full Generalized Procrustes Analysis (full GPA; *procGPA* function of shapes package version 1.2.3 for R version 3.4.3 in RStudio version 1.1.383; Dryden, 2017; R Core Team, 2017; RStudio Team, 2017). The transformations applied to each mean landmark configuration were determined by superimposing each untransformed mean configuration onto the corresponding transformed configuration. Centroid coordinates and centroid size were computed (Zelditch et al. 2012) and used to determine translation and scaling parameters. The rotation matrix was computed as the matrix UV^T, where V and U were the left and right matrices of the singular value decomposition VTU^T of the product of the transposed transformed and untransformed mean landmark matrices (Dryden & Mardia, 2016, lemma 4.2). The transformations of the mean configurations were applied to all corresponding landmark sets. This allowed for a full Procrustes superimposition among the means of the separate girdle halves of the specimens while preserving the variance (measurement error) within a given girdle half for further analysis.

The transformed landmark sets of Landmark Dataset 1 were uniformly rescaled such that the centroid size of the mean landmark configuration of the ventral girdle half of ZMH A12601 after superimposition equaled the corresponding centroid size before superimposition ('superimposed Landmark Dataset 1'). This allowed for a maximum comparability to Landmark Datasets 2 and 3. All computations were performed using basic R functions and functions of the packages geomorph (version 3.0.5; Adams et al. 2017), abind (version 1.4-5; Plate & Heiberger, 2016), and shapes. Landmark Datasets 2, 3, and 4 were analysed without any superimposition as the analyses were performed for each scan/specimen separately.

Visualization

Landmark locations were visualized for the ventral girdle half of *B. orientalis* specimen ZMH A12601 (NoDown, MidGreyT, subSimpl) in MODO® (version 10.1v2; The Foundry) by

creating spheres with midpoint coordinates adopted from an arbitrary landmark set acquired from that surface. Deviations among the 'reasonable' surface variants of selected specimens (ventral girdle halves of ZMH A05110, A05619, A05681, A09674, A12601) were visualized in GOM Inspect 2017 (GOM GmbH). Local deviations were calculated and color-coded per vertex (Fig. 1); simplified surfaces had a low vertex count and their vertex count was increased prior to visualization (MODO®, *Subdivide: Faceted* function) for higher resolution distance mapping in GOM.

Principal component analyses were performed for Landmark Datasets 1–3 separately in R. The results were visualized using the packages ggplot2 (Wickham, 2016) and ggpubr (version 0.1.6; Kassambara, 2017). All figures were arranged in Adobe® Illustrator® CS6 (version 16.0.3; Adobe® Systems Software).

Permutational analyses of variance

Permutational MANOVAs (PERMANOVAs; Anderson, 2001) were performed for each of (superimposed) Landmark Datasets 1–3, treating all factors as random and nested (Arnqvist & Mårtensson, 1998). In particular, for the superimposed Landmark Dataset 1, the residual term reflected the variance of the repetitions by each observer and was nested within observer. Observer was nested within simplification, which was nested within segmentation, which was nested within position, which was nested within specimen, which was nested within species (compare Table 2). For Landmark Dataset 2, the residual term reflected the variance of the repetitions of the same day and was nested within the factor day. Day was nested within simplification, which was nested within segmentation, which was nested within downsampling (compare Table 3). For Landmark Dataset 3, the residual term reflected the variance of the repetitions of the same day and thus was nested within day; day was nested within observer. P-values < 0.05 were considered significant for all tests.

PERMANOVAs were computed in R using the *adonis* function of the vegan package (version 2.4-5; Oksanen et al. 2017) with Euclidean distance as distance measure and 9999 permutations. Permutations were performed for each factor separately using the mean configurations of the nested groups defined by the respective next-lower factor (ensuring correct computation of F-values) and restricted within the groups defined by the next-higher factor (compare Anderson & ter Braak, 2003). Permutations of the lowest factor were performed on the landmark sets; permutations of the highest factor were unrestricted. To test, for example, the significance of the factor specimen in the superimposed Landmark Dataset 1, the average landmark configurations of each girdle half were computed for each specimen (in

other words, the mean configurations of the groups defined by the next-lower factor position nested within specimen). Permutations were restricted within species and the reduced model comprised the factors 'specimen' nested within species; if there had been higher factors, those would have been part of the reduced model as well. The corresponding *F*- and *P*-values of the full model (correct values for degrees of freedom, sum of squares, and mean squares) were replaced by those of the reduced models. The relative contribution of each factor to the total variance (variance components expressed as percentages) in a respective landmark dataset was computed following Sokal & Rohlf (1981); negative variance components were set to zero and not considered for percentage calculation (only applicable for the factor simplification in Landmark Dataset 1).

Using different scanner types and software packages for volume data acquisition might have induced artefactual variance in the superimposed Landmark Dataset 1; we did not account for this, as a previous study found only minor effects of scanner type on the geometry of derived polygon surfaces (Colman et al. 2017). If there was artefactual variance caused by CT scanning and volume reconstruction, this would be incorporated in the factor 'specimen' in the PERMANOVA. Consequently, the measured variance due to specimen would artefactually be higher than it actually was and measurement error would be slightly underestimated.

Variance within and between subgroups

Selected sets of pairwise Euclidean distances between landmark configurations were computed to compare informally the magnitudes of the variation within different subgroups. The distance sets were individually pooled according to selected factors (i.e. distances within sub-groups were treated as one distance set if they were associated with the same level of a given factor) to assess informally the dependence of variance within subgroups on the different levels of the factors. A small variance, and thus an overall high precision in landmark placing, became apparent in short pairwise distances within the (pooled) subgroups. Notched boxplots (McGill et al. 1978) of the (pooled) groups of distances were used to identify substantial differences among groups. All calculations were performed in R.

In particular, for superimposed Landmark Dataset 1, the pairwise Euclidean distances of the 18 landmark sets acquired from the same girdle half by the same observer were calculated for each observer separately resulting in 153 distances per observer per girdle half. The distances were pooled by specimens; in particular, all distances of a given specimen, calculated within observer and girdle halves separately, were treated as one set of distances.

This allowed for an informal assessment of the precision with which the surfaces of a given scan could be landmarked. For Landmark Dataset 2, pairwise distances were computed among the nine landmark sets acquired from the same surface variant, resulting in 45 sets of 36 distances each. The distance sets were pooled according to the levels of downsampling, segmentation, and simplification, respectively, to assess informally the dependence of the overall landmark precision on the respective levels of the factors. Similarly, for Landmark Dataset 4, pairwise distances were computed among the nine landmark sets acquired from the same surface variant.

Landmark precision and observer differences

All landmark sets in Landmark Dataset 3 were acquired from the same surface and thus were superimposed without any GPA. This allowed for the direct analysis of the precision with which the landmarks could be placed. Therefore, the standard deviation of each landmark in Landmark Dataset 3 was calculated according to von Cramon-Taubadel et al. (2007) for each observer and each day separately to measure the precision with which a given landmark could be placed. Separate Wilcoxon signed rank tests were performed for each observer to test for significant differences in the landmark standard deviations between days. The Euclidean distances of each landmark configuration in Landmark Dataset 3 from the mean configuration obtained by O1 (O1 trained all observers, therefore the mean configuration of O1 was set as reference) were calculated to assess the similarity of the shapes measured by the different observers to the reference. The consideration of the day and order, in which the landmark sets were acquired, allowed for assessing trends in potential systematic deviation from the reference by linear regression. Calculations and visualizations were done in R using the above-mentioned packages and plotrix (Lemon, 2006).

Results

‘Reasonable’ surface variants and true biological variation

A visual comparison of the ‘reasonable’ surface variants (MidGreyT, OtsuT, SubThresh; original, subSimpl) of the ventral pectoral girdle halves of selected specimens (see Fig. 1 for ZMH A12601) revealed that surfaces of a given specimen rarely differed by more than two voxels; the highest deviations mainly occurred in areas where no landmarks had been placed. Subjective optimal surface simplification (subSimpl) seemed to have a smaller effect than one voxel. This indicates that the simplification removed the voxel-steps from the surfaces while

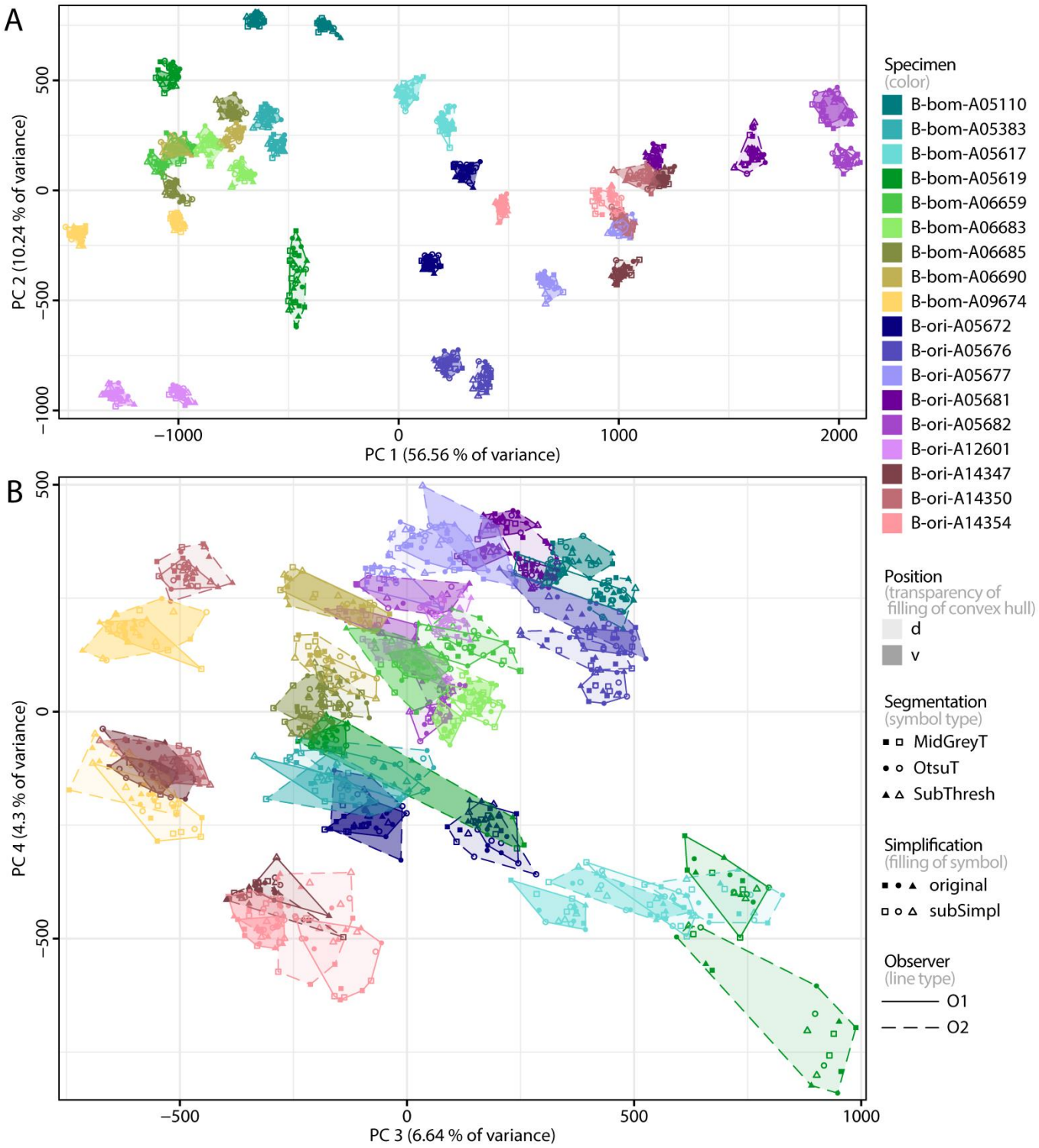


Fig. 3 Plots of principal components of the landmark sets acquired from the 'reasonable' surface variants (Landmark Dataset 1). Convex hulls encircle all landmark sets of a given girdle half acquired by the same observer. Specimens denoted by color, position by transparency of filling of the convex hull, segmentation by the type of the symbol, simplification by the filling of the symbol, and observer by the type of the line used for the convex hull. (A) Principal components 1 and 2. (B) Principal components 3 and 4.

maintaining the gross geometry and the anatomical details (e.g. no artefactual deformation of edges).

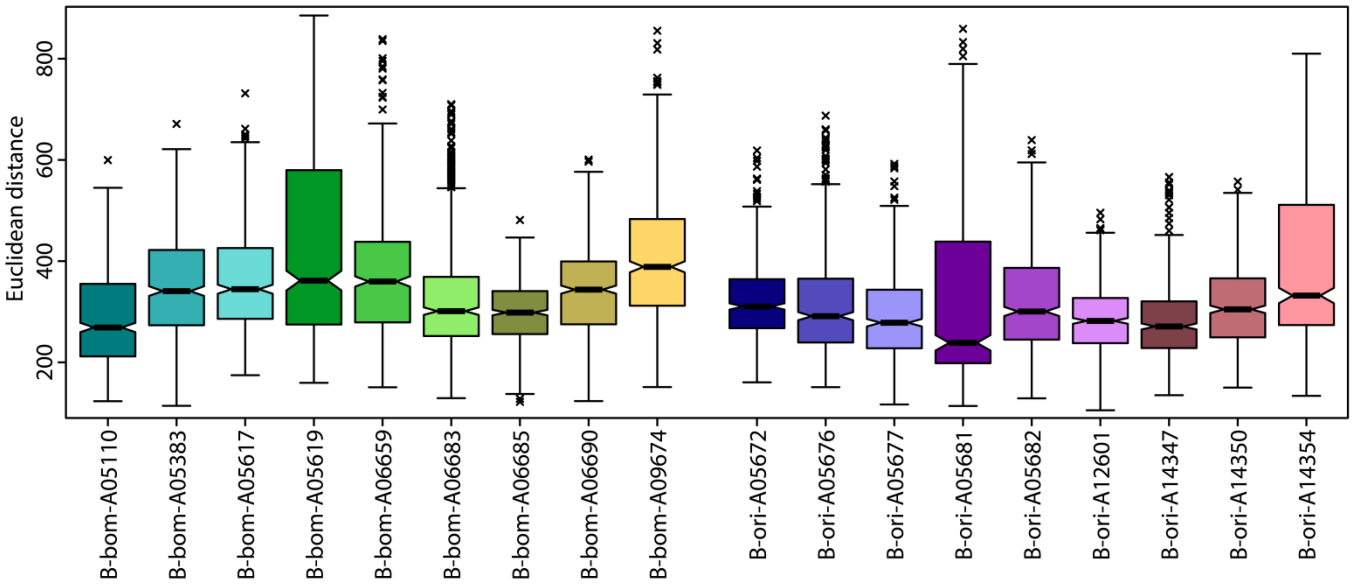


Fig. 4 Boxplots of pairwise distances between landmark sets of a given girdle half acquired by the same observer pooled by specimens (used as informal measure of overall landmark precision by specimens).

A plot of the first two principal components (56.56 and 10.24% of total variance, respectively) of the superimposed landmark sets acquired from the 'reasonable' surface variants (superimposed Landmark Dataset 1; Fig. 3A) showed that the sets from a given girdle half generally clustered together irrespective of surface variant and observer. Most girdle halves were separated along principal components 1 and 2. The specimens with overlapping regions in the first two principal components were separated along the third through fifth (6.64, 4.3, and 4.08%, respectively; latter not shown; Fig. 3B) principal components. Boxplots (Fig. 4) of pairwise Euclidean distances between the landmark sets acquired by a given observer from a given girdle half pooled by specimens showed considerable differences in the overall variance in the landmark data among specimens.

A nested PERMANOVA of superimposed Landmark Dataset 1 revealed significant contributions of the factors species, specimen, position, segmentation, and observer to the total variance in the landmark data (Table 4), with specimen being the major factor, accounting for 47.92% of total variance. 93.25% of the total variance was caused by true biological variation. The major factor causing artefactual variance was intra-observer error with a contribution of 3.1% to the total variance, followed by inter-observer error, which contributed 2.86%. Segmentation accounted for 0.79% and the factor simplification was not significant.

Table 4 Nested PERMANOVA of the landmark sets acquired from 'reasonable' surface variants of different specimens (superimposed Landmark Dataset 1).

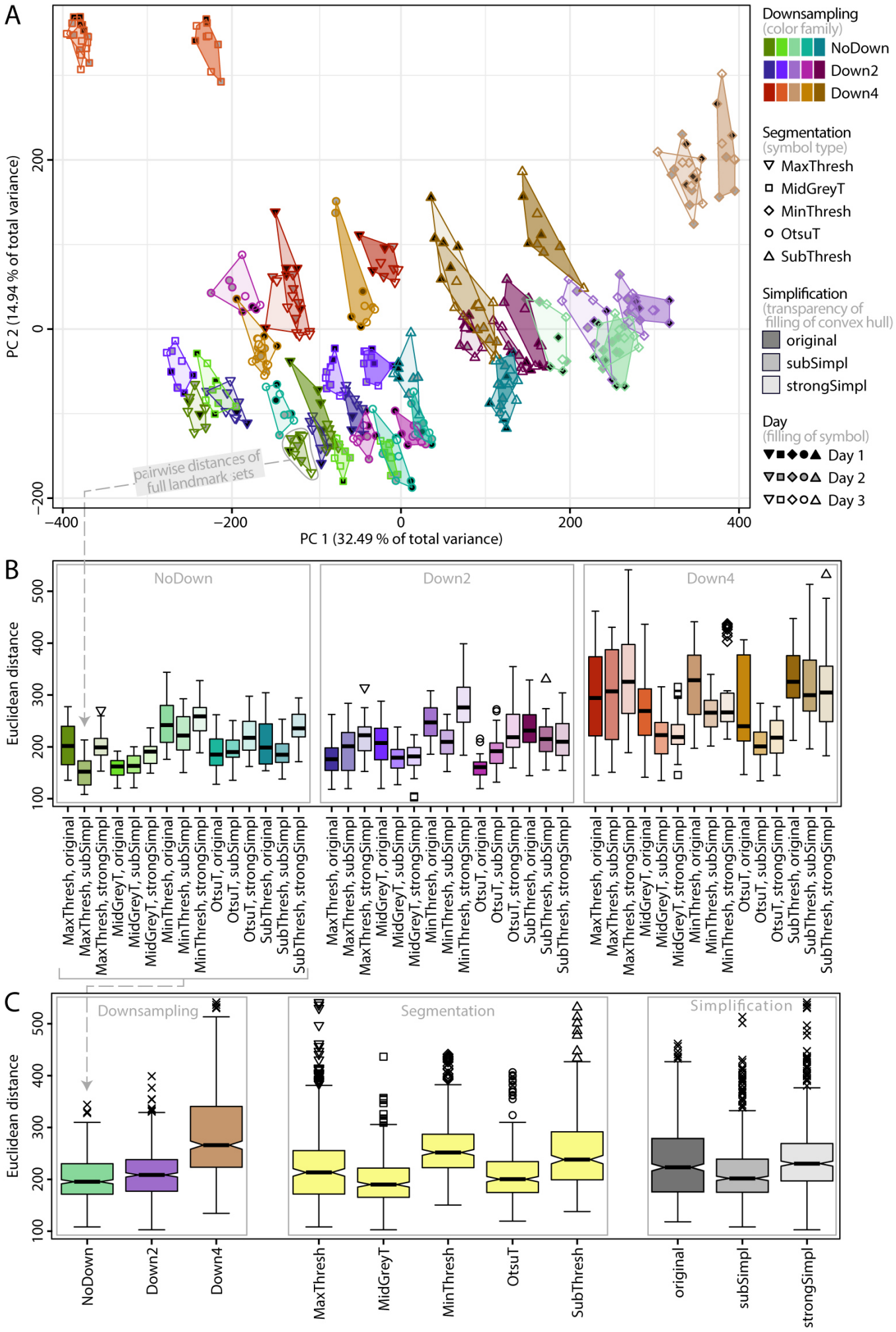
	df	**SS**	**MS**	***F***	***P***	**Variance component [%]**
Species	1	746575982	746575981.54	10.26	0.0007***	31.37
Specimen (nested in species)	16	1164604487	72787780.41	7.71	0.0001***	47.92
Position (nested in specimen)	18	169950185	9441676.92	44.06	0.0001***	13.96
Segmentation (nested in position)	72	15427585	214272.02	5.30	0.0001***	0.79
Simplification (nested in segmentation)	108	4373945	40499.49	0.19	1	0
Observer (nested in simplification)	216	46331768	214498.93	3.77	0.0001***	2.86
Residuals (repetitions nested in observer)	864	49175843	56916.48			3.10
Total	1295	2196439793				

All factors treated as random; permutations (if applicable, of means of next-lower factor) performed for each factor separately and, if necessary, restricted within groups of next-higher factor. *** $P \leq 0.001$.

Maximum range of surface variants

Figure 5A shows a plot of the first two principal components of the landmark sets that were acquired from the maximum range of surfaces of a selected girdle half (Landmark Dataset 2). Principal components 1 and 2 represented, respectively, 32.49 and 14.94% of the total variance in the landmark data. Landmark sets from surfaces generated with different segmentation strategies and different degrees of surface simplification were roughly separated along the first principal component, and the strongest downsampling degree (Down4) was roughly separated from the other two (NoDown, Down2) along the second and third (8.64% of total variance, not shown) principal components. There was no obvious pattern of clustering in plots of higher principal components (not shown), which indicates a ratherandom variation of landmark data along these components. However, not all groups of landmark sets of distinct surface variants were perfectly separated. This potentially indicated considerable similarity of the corresponding surfaces and was particularly true for subjectively optimal and strongly simplified (subSimpl, strongSimpl) surface variants of the strongly downsampled (Down4) volume. T

he boxplots of the Euclidean distances derived from Landmark Dataset 2 (Fig. 5B,C) generally showed highest variations (greatest pairwise distances) for the landmark sets acquired from surfaces of strongly downsampled (Down4) volumes compared with the other two degrees of downsamling (NoDown, Down2). Within the factor segmentation and irrespective of the other factors, the landmark sets acquired from surfaces generated with the segmentation strategies MidGreyT and OtsuT, on average, showed the least variance. With regard to surface simplification, landmark sets on subjectively optimal simplified (SubThresh)


A
PC 2 (14.94 % of total variance)
PC 1 (32.49 % of total variance)
pairwise distances of full landmark sets
Downsampling (color family)
NoDown
Down2
Down4
Segmentation (symbol type)
MaxThresh
MidGreyT
MinThresh
OtsuT
SubThresh
Simplification (transparency of filling of convex hull)
original
subSimpl
strongSimpl
Day (filling of symbol)
Day 1
Day 2
Day 3
B
Euclidean distance
NoDown
Down2
Down4
MaxThresh, original
MaxThresh, subSimpl
MaxThresh, strongSimpl
MidGreyT, original
MidGreyT, subSimpl
MidGreyT, strongSimpl
MinThresh, original
MinThresh, subSimpl
MinThresh, strongSimpl
OtsuT, original
OtsuT, subSimpl
OtsuT, strongSimpl
SubThresh, original
SubThresh, subSimpl
SubThresh, strongSimpl
C
Euclidean distance
Downsampling
Segmentation
Simplification
NoDown
Down2
Down4
MaxThresh
MidGreyT
MinThresh
OtsuT
SubThresh
original
subSimpl
strongSimpl

← **Fig. 5** Principal component plot and Euclidean distances of landmark sets acquired from the maximum range of surface variants of the ventral pectoral girdle half of *Bombina orientalis* (ZMH A12601; Landmark Dataset 2). (A) Plot of first two principal components. Convex hulls encircle all landmark sets of a given surface variant. Downsampling denoted by color family, segmentation by the type of the symbol, simplification by transparency of filling of the convex hull, and day by the filling of the symbol. (B) Boxplots of pairwise Euclidean distances between the full landmark configurations of each surface variant. (C) Notched boxplots of pairwise Euclidean distances of (B) pooled according to the levels of, respectively, the factors downsampling, segmentation, and simplification. (Pooled) Euclidean distances were used to informally compare the variance among the levels of the factors.

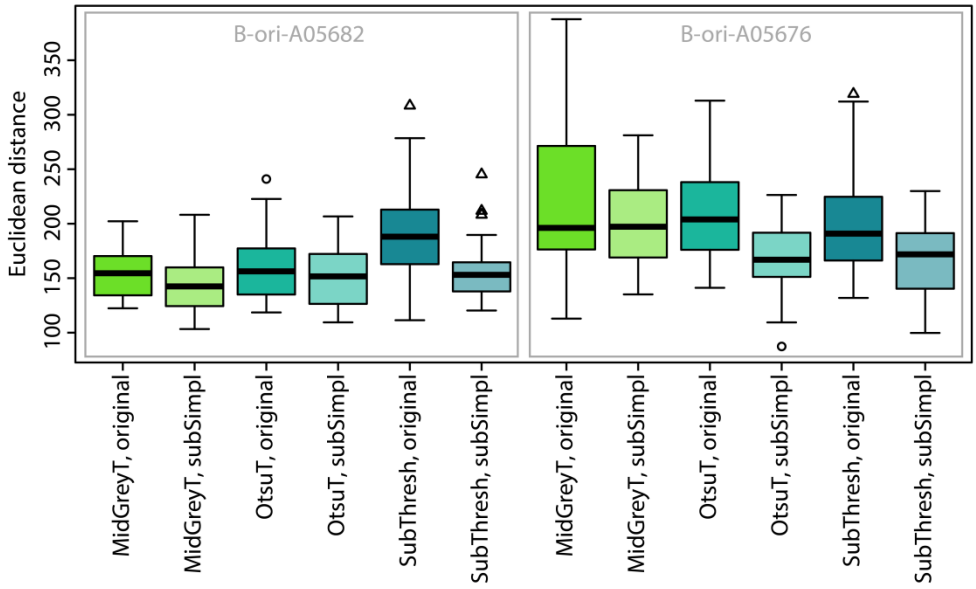


Fig. 6 Boxplots of pairwise Euclidean distances between the landmark sets acquired from each 'reasonable' surface variant of *Bombina orientalis* specimens ZMH A05682 (left) and ZMH A05676 (right).

surfaces generally were the least variable. This was in agreement wi th the subjective impression of O1 during landmark acquisition: landmark placing on subjective optimal simplified surfaces was experienced as being easiest (also applies for other scans/specimens).

The pattern of Euclidean distances within the landmark sets of the surface variants of *B. orientalis* specimen ZMH A05682 in Landmark Dataset 4 ('reasonable' surface variants of ZMH A05676 and A05682, each repeatedly landmarked three times in three sessions; Fig. 6) was similar to that of ZMH A12601 in Landmark Dataset 2 (Fig. 5B): segmentation based on automatic local thresholding (MidGreyT, OtsuT) and subjectively optimal surface simplification (subSimpl) was advantageous with regard to overall landmark precision. For ZMH A05676 (Fig. 6), landmark sets from surface variants derived from automatic local thresholding based segmentations showed similar (OtsuT) or higher (MidGreyT) variation than those sets acquired from the surfaces created by manual threshold selection (SubThresh).

Table 5 Nested PERMANOVA of the landmark sets acquired from the maximum range of surface variants of the ventral (superficial) girdle half of *Bombina orientalis* (ZMH A12601; Landmark Dataset 2).

	df	**SS**	**MS**	***F***	***P***	**Variance component [%]**
Downsampling	2	5745432	2872715.82	1.5201	0.1397	6.14
Segmentation (nested in downsampling)	12	22678045	1889837.10	9.0826	0.0001***	52.56
Simplification (nested in segmentation)	30	6242139	208071.28	4.5536	0.0001***	15.22
Day (nested in simplification)	90	4112486	45694.29	1.9432	0.0001***	6.24
Residuals (repetitions nested in day)	270	6349050	23515.00			19.84
Total	404	45127151				

All factors treated as random; permutations (if applicable, of means of next-lower factor) performed for each factor separately and, if necessary, restricted within groups of next-higher factor. *** $P \leq 0.001$.

The variance among landmark sets acquired from subjectively optimal simplified (subSimpl) surfaces of ZMH A05676 generally was smaller than the variance among the respective original surfaces.

The nested PERMANOVA of Landmark Dataset 2 revealed a significant contribution of the factors segmentation, simplification, and day to the total variance within the landmark data. Among those factors, segmentation (52.56%) was responsible for most variation (Table 5). The variance between days was smaller than the variance within the same day and also smaller than the added variance due to surface simplification. Downsampling did not contribute significantly to the total variance.

Observer error and landmark precision

A plot of the first two principal components (33.26 and 28.08% of total variance, respectively) of Landmark Dataset 3 (different observers on same surface, 20 repetitions on each of two days, first five repetitions of each day discarded) showed that the landmark sets of a given observer generally cluster together (Fig. 7A), with overlapping areas between O1 and O3, as well as between O5 and O6. The Euclidean distance of each landmark set to the mean configuration of O1 (set as reference) revealed that the deviations of O2, O3, O5, and O6 were more or less constant over time (Fig. 7B). The deviations of O4 initially fell in the same range as those of O2, O3, O5, and O6, but they increased with time.

A nested PERMANOVA of Landmark Dataset 3 revealed that the factors observer and day significantly contributed to the total variance (70.36 and 5.53%, respectively; Table 6). The variation within the repetitions of a given day contributed 24.1%.

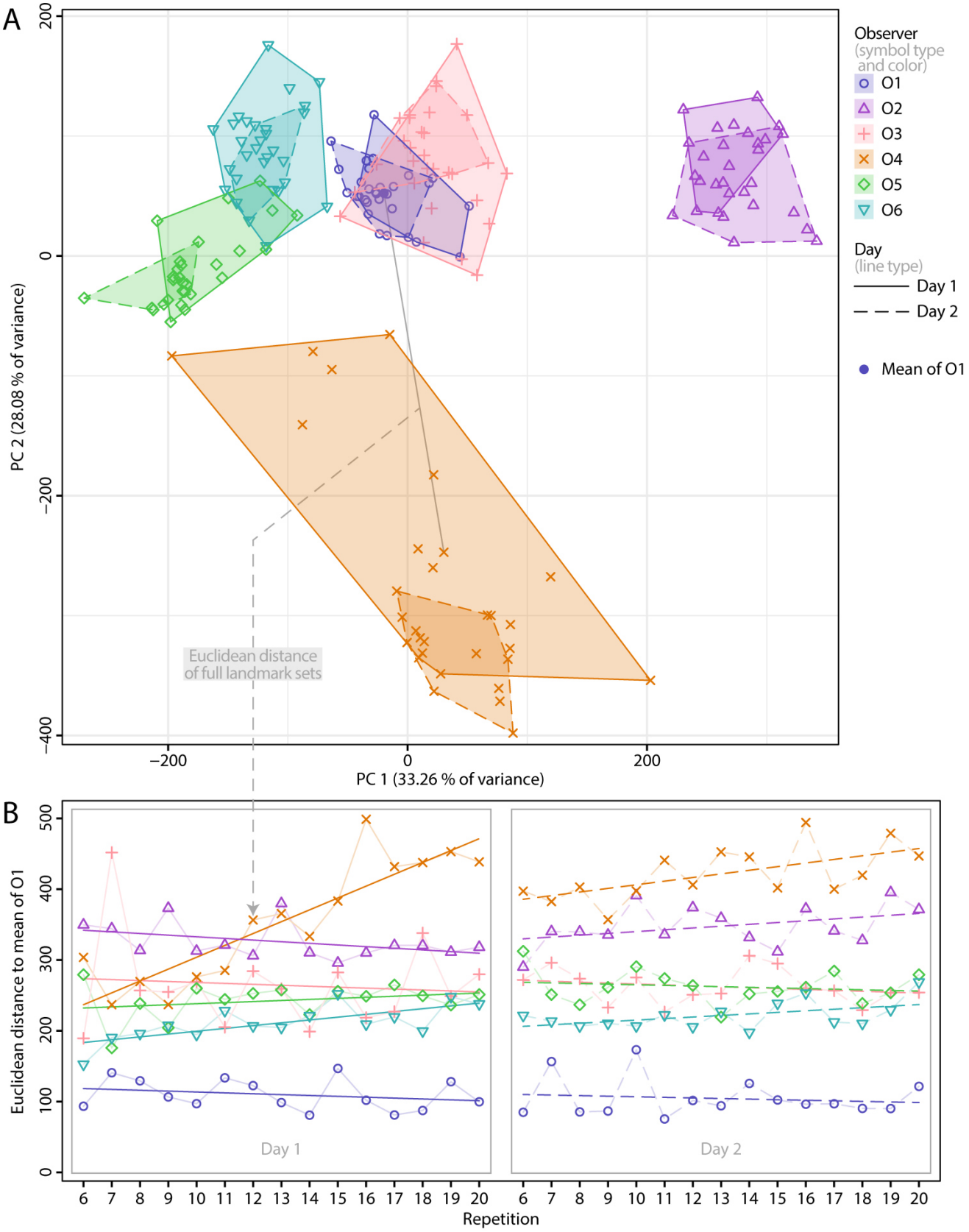


Fig. 7 Principal component plot and Euclidean distances of landmark sets to the mean landmark configuration of O1 derived from Landmark Dataset 3 (six observers, one surface variant). (A) Principal components 1 and 2. Convex hulls encircle all landmark sets by a given given observer acquired on the same day. Observer denoted by symbol type and color, and day by the type of the line used for the convex hull. (B) Euclidean distances of the full landmark sets to the mean landmark configuration of O1 (set as reference) by day and order of acquisition. Regression lines visualize trends in deviation from the reference.

Table 6 Nested PERMANOVA of the landmark sets acquired by different observers from the same surface variant (MidGreyT, original) of ventral (superficial) girdle half of *Bombina orientalis* (ZMH A12601; Landmark Dataset 3).

	df	**SS**	**MS**	***F***	***P***	**Variance component [%]**
Observer	5	8432348	1686469.66	20.71	0.0002***	70.36
Day (nested in observer)	6	488501	81416.87	4.44	0.0001***	5.53
Residuals (repetitions nested in day)	168	3078948	18327.07			24.10
Total	179	11999797				

All factors treated as random; permutations (if applicable, of means of next-lower factor) performed for each factor separately and, if necessary, restricted within groups of next-higher factor. *** $P \leq 0.001$.

The smallest standard deviation across observers was below 10 for all landmarks, with the exception of landmarks 4 and 20 (10.11 and 11.99, respectively; Fig. 8). Some landmarks (e.g. 2, 11, 24) were acquired with consistent high precision by all observers, whereas the standard deviations of others (e.g. 7, 18, 20) greatly differed among observes. Wilcoxon signed rank tests revealed no significant differences in the standard deviations of the different days for O1 and O2 (*P*-values 0.2522 and 0.1974, respectively). For O3– O6, the standard deviations significantly differed among days (0.0031, 0.0164, 0.0007, and 0.0005, respectively); the standard deviations of the second day were generally smaller, but exceptions occurred.

Discussion

In µCT-based 3D geometric morphometrics, data goes through several processing steps, each of which may add artefactual variance to the final landmark data. Researchers often follow commonly used procedures and protocols without full quantitative appreciation of measurement error that is potentially introduced with each of the processing steps. We intended to assess the artefactual variance that had been added during surface generation and landmark acquisition to see whether some of the steps are more critical than others, and to derive recommendations for measurement error reduction. We identified variance introduced by observer and segmentation as the main sources of measurement error. Training periods prior to landmark acquisition, landmark acquisition in as few sessions as possible, careful surface simplification, and the use of case-specific optimal segmentation strategies can potentially help reduce measurement error.

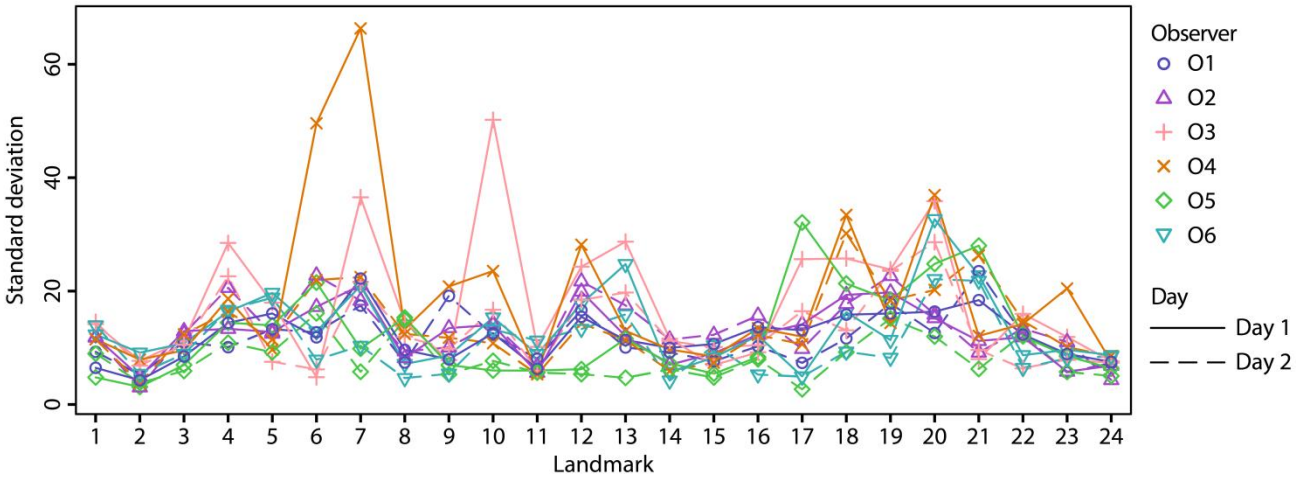


Fig. 8 Landmark standard deviations calculated for each observer and each day separately (derived from Landmark Dataset 3).

Quality of automatic local thresholding

Our choice of automatic local thresholding algorithms and corresponding parameters was based on previous studies (*Otsu* algorithm; Landini et al. 2017; Healy et al. 2018) and on the performance of different algorithm-parameter combinations applied to a rather arbitrarily created stack of reconstructed phantom images (*MidGrey* algorithm). We did not evaluate the effects of image noise, pixel size or contrast of bone and soft tissue. These factors might have a significant effect on the thresholding quality. Furthermore, various other thresholding methods have been published (reviewed, e.g. by Sezgin & Sankur, 2004); these were not considered herein due to the limitation to the use of the *Auto Local Threshold* plugin. It is likely that the algorithms and parameters, as well as the strategy of thresholding resliced versions of the µCT stacks and combining them later on, may not have led to optimal results for all µCT scans. We believe, however, that our automatic local thresholding strategies yielded good results, as a visual control showed acceptable surfaces; our aim was to cover a range of surfaces to assess the effect of different segmentation strategies in CT-based geometric morphometric studies, not to rate the performance of different automatic thresholding strategies. Still the questions remain, which automatic thresholding strategy results in a binarization closest to reality, and how the choice of the optimal thresholding strategy depends on, for example, the scan quality. Answering these questions will allow avoidance of measurement error caused by using unnaturally shaped surface geometries for landmark acquisition.

Observer error and surface simplification

Measurement error caused by artefactual variance between and within observers is a common phenomenon in geometric morphometric studies (e.g. Valeri et al. 1998; Barbeito-Andrés et al. 2012). The data reported in the present study are no exception: inter- and intra-observer errors were the major source of artefactual variance in landmark data that represented a realistic geometric morphometric dataset (superimposed Landmark Dataset 1; Table 4). In the superimposed Landmark Dataset 1 the relative contributions of inter- and intra-observer error were similar, the latter only slightly exceeding the former. This observation contradicts previous findings that inter-observer error commonly exceeds intra-observer error (e.g. Singleton, 2002; Wilson et al. 2011). The pattern observed in our data might be caused by the composition of Landmark Dataset 1: repetitions on a given surface were performed on different days and all other surfaces were landmarked in-between repetitions. This might have prevented any effect of memorizing exact landmark positions from the previous repetition. In addition, the number of surfaces landmarked consecutively might have caused some kind of fatigue which, in turn, might have led to inattentive landmark placing, causing higher intra-observer error. Further, inter-observer error in Landmark Dataset 1 might have been exceptionally small, as both observers extensively discussed landmark positions prior to landmark acquisition. The pattern observed in Landmark Dataset 3, acquired by six partly inexperienced observers, agrees well with other reports, as inter-observer error has clearly exceeded intra-observer error (Table 6). The observed relatively large Euclidean distances of O2 to the mean landmark configuration of O1 (reference) and the increasing distances of the landmark sets acquired by O4 (Fig. 7) indicate systematic deviations. These suggest that there were differences in how observers identified landmarks (i.e. they did not identify exact homologous points as the spots where to place the landmarks) and that they have placed the landmarks at different points. The particular pattern observed for O4 might also have been attributed to O4 having a ‘bad day’, leading to inconsistent landmark placing. Considering, however, that the deviations of O4 seem to have a direction, having a ‘bad day’ seems to be an unlikely explanation for the observed pattern. Inconsistency in landmark identification even among experienced observers has also been reported by Shearer et al. (2017) and seems to be a common phenomenon. This should caution researchers to make sure that landmarks are precisely defined and that all observers place landmarks at exactly homologous points; in-person training in landmark collection can help minimizing inter-observer error (Shearer et al. 2017). The observers should be trained to have a thorough knowledge of the landmark acquisition techniques, as well as of the biological landmarks and the variability in their

expressions (Corner at al. 1992). Taking into account that there are deviations even between experienced observers, it seems advisable to test repeatedly for inter-observer inconsistencies in landmark identification over time to prevent observers from systematically deviating from the defined landmarks by developing their own landmark definitions.

Despite the considerable amount of inter-observer error observed in previous works and herein, previous accounts (e.g. Singleton, 2002; Chang & Alfaro, 2016) and the fact that all specimens in our superimposed Landmark Dataset 1 were well-separated along the first few principal components, show that landmark data acquired by different observes can give results that are precise enough to allow correct biological inferences. If measurement error is crucial, that is, if the biological variation of interest is small relative to error, as for example in the analysis of asymmetry (Klingenberg, 2015; Robinson & Terhune, 2017; Shearer et al. 2017), it seems advisable to exclude inter-observer error by having landmark acquisition be performed by only one observer. Other strategies to cope with inter-observer error have been suggested in literature (see Fruciano, 2016 and Fruciano et al. 2017 for a more detailed discussion).

Intra-observer error contributed considerably to measurement error in parts of the present study. A common suggestion in the literature is to reduce intra-observer error by performing repeated measurements of landmarks and to use them or their averages for analyses (Corner et al. 1992; Arnqvist & Mårtensson, 1998; also see review by Fruciano, 2016). In our study, the considerable decrease in landmark standard deviations on the second day exclusively in inexperienced observers (Fig. 8) implies an increase in the precision of landmark placing with experience. A similar learning effect has been reported by Valeri et al. (1998). As with inter-observer error reduction, including a training period seems to be an effective way of decreasing intra-observer error, too (also see Chang &Alfaro, 2016).

Our data suggest that explicitly using surfaces that allow for a high precision in landmark placing is an additional way to reduce intra-observer error. The relatively small Euclidean distances between repeated landmark sets of a given subjectively optimal simplified (subSimpl) surface variant imply that a slight surface simplification generally increases the precision with which landmark coordinates can be acquired (Figs 5B,C and 6), whereas the alteration of the surface geometry seems negligible (Fig. 1. Table 4). With stronger simplification, however, negative effects set in and increase the artefactual variance (Fig. 5B,C, Table 5) by altering more and more the surface geometry and by decreasing the landmark placing precision in most cases. Generally, we recommend the use of surface simplification, but this has to be applied with caution; its effect in a given study should be

assessed appropriately. Our data suggest that another way of obtaining surfaces that allow for a high precision in landmark acquisition might be the application of wisely chosen segmentation strategies; segmentation based on automatic local thresholding using case-specific optimal combinations of thresholding algorithms and parameters seems to outperform manual thresholding (Figs 5B,C, 6, and see below).

Error within and between days

In Landmark Datasets 2 and 3, the artefactual variance within the repetitions on the same day exceeded the variance between days (Tables 5 and 6). This observation is counterintuitive in that, when repeating landmark acquisition on the same day and on different days, one would expect that, due to short-term memory effects, the measurement error between days would be greater than within repetitions on the same day. All landmark sets in Landmark Dataset 2 were acquired from different surface variants of the same specimen and O1 knew the anatomical peculiarities of that specimen very well. As a consequence of this, it seems likely that O1 remembered the exact landmark positions the next day, which might have led to unusually small variance between days. The high number of consecutive repetitions on the same day might have led to inattentive landmark placing which caused higher measurement error within days. The observed pattern of observer-dependent variance in Landmark Dataset 3 might have similar causes. Additionally, it seems likely that, for the inexperienced observers, it took several repetitions to find their own way of identifying landmarks; this would have increased the observer error on the first day.

Error caused by segmentation

Segmentation based on automatic local thresholding algorithms (MidGreyT, OtsuT) outperformed manual thresholding in the two *B. orientalis* specimens ZMH A05682 and ZMH A12601 (ventral girdle halves only; Figs 5B,C and 6), for which the combination of thresholding algorithm and parameters likely was close to optimal. For these two specimens, automatic local thresholding had two advantages: first, the derived surfaces most likely had a geometry closer to the real bones than did the surfaces of other segmentation strategies and, second, the generated surfaces allowed for placing landmarks with higher precision and thereby helped reducing intra-observer error. These positive effects, however, did not apply to all specimens. Among the surfaces of ZMH A05676 (ventral girdle half only), the surface derived by manual threshold selection (SubThresh) allowed for equal or higher precision in landmark placing and, thus, outperformed some of the automatic-local-thresholding-derived

surfaces. One explanation might be that the quality (CNR of bone and soft tissue) of the µCT scan of ZMH A05676 was quite different from those of ZMH A05682 and ZMH A12601 and the parameters used during automatic local thresholding were less optimal for the scan of ZMH A05676. This might have resulted in a specific surface geometry less suitable for precise landmark placing. In other words, the application of the case-specific optimal combination of automatic local thresholding algorithm and parameters can outperform other segmentation strategies with regard to measurement error, whereas non-optimal combinations can increase measurement error. These conclusions are based on only three specimens and need to be verified in future studies. Yet, segmentation had a significant effect on measurement error in our data. Our findings corroborate previous studies (Williams & Richtsmeier, 2003; Gunz et al. 2012) that have reported thresholding to be a critical step when deriving measurements from CT-based surfaces. This evidence suggests that special attention should be paid during the selection of the thresholding strategy in geometric morphometric studies. Ideally, the effects of using different segmentation strategies should be assessed; yet, considering that only 0.79 % of variance in Landmark Dataset 1 was caused by the factor segmentation, a formal comparison of different thresholding strategies might not be mandatory. If automatic thresholding algorithms were applied, the use of case-specific optimal algorithm-parameter combinations should be assured to prevent negative effects (i.e. increased measurement error).

Effect of voxel size

Our statistical analyses suggest that the resolution (i.e. the voxel size) of the underlying µCT-derived volume does not significantly contribute to measurement error in landmark data (Table 5). The patterns in Fig. 5, however, indicate a considerable impact of resolution on shape measured by landmarks for surfaces derived from strongly downsampled (Down4) volumes. Downsampling had two effects. First, the more volumes were downsampled, the more the shapes measured from the derived surfaces deviated from the shape of the corresponding surface generated from the full resolution volume (NoDown; Fig. 5A). Second, landmark sets acquired from surfaces of strongly downsampled volumes were considerably more variable (higher intra-observer error) than those of less downsampled volumes (Fig. 5B,C). The insignificance of the factor downsampling in the PERMANOVA of Landmark Dataset 3 (Table 5) might have been caused by the small number of repetitions; more surface variants of additional downsampling degrees or the addition of differently downsampled scans of other specimens might have resulted in significance. One potential reason for the peculiar

pattern observed in Fig. 5A is the occurrence of non-random variance due to downsampling. Such non-random variance might have been too small to cause significance in the PERMANOVA but might have caused the rough separation of landmark sets according to the degree of downsampling along principal component 2.

Christiansen (2016) has reported a strong effect of voxel size on measures of trabecular bones if structures were too thin relative to voxel size. Simon & Marroig (2015), however, found no considerable difference in the precision of landmarks from volumes with a range of different voxel sizes. The evidence from previous studies and the comparable small differences between landmark sets from surfaces of non-downsampled (NoDown) and corresponding slightly downsampled (Down2; Fig. 5A) volumes suggest that the voxel size has a minor effect on measurement error, given that it is small enough relative to the structures examined. The use of volume data with a coarse spatial resolution in relation to structure size may cause considerable artefactual variance. Thus, shape analysis of specimens of similar sizes scanned with moderately different voxel sizes might be uncritical as long as the spatial resolution is small enough. Although the generality of these conclusions remains to be proven, it still seems advisable to use volume data with a good spatial resolution to avoid any potential increase of measurement error. On the other hand, operating a CT scanner at the highest possible spatial resolution considerably increases scanning and image processing times and can cause lower signal-to-noise ratios in the X-ray projections that are captured during a scan. Therefore, finding a reasonable case-specific resolution is recommended.

With regard to our methodological approach for simulating different scan resolutions, it should be noted that the downsampling (Down2, Down4) might have decreased image noise in the volume data. As a consequence, the respective segmentation results and thus the derived surfaces might have been of a better quality than those of the non-downsampled (NoDown) data. The variance components for the factors segmentation, simplification, and day, as well as for the repetitions within one day (residuals), might slightly underestimate the true error (Table 5). The decisions on the significance of the respective factors should not be affected by this, as all those factors are already highly significant, even with the potentially underestimated artefactual variance.

Landmark precision

The pattern of landmarks standard deviations observed in Landmark Dataset 3 (Fig. 8) shows that all landmarks can be placed with high precision, which indicates that the definition of landmarks itself is sufficiently precise. The higher standard deviations observed for some

landmarks might indicate a personal component in landmark identification and placing precision; that is, some observers have difficulties in placing certain landmarks, whereas others are able to place these landmarks with high precision. For dealing with such error-prone landmarks, von Cramon-Taubadel et al. (2007) suggested either excluding the landmarks concerned from the dataset or redefining them more accurately. It is remarkable, however, that the highest standard deviations in our data occurred on the first day and were produced by inexperienced observers. Thus, the imprecision in landmark placing in Landmarks Dataset 3 might mainly be the result of lack of experience.

The overall precision of landmarks varied considerably among specimens (Fig. 4). The lack of a clear pattern of landmark precision correlating with scan parameters or bone to soft tissue CNR (Table 1), and the lack of major differences in the variation among surface variants of a given specimen, suggest that specimen morphology has caused the differences in landmark precisions. This implies different levels of measurement error among individual specimens. Therefore, it seems advisable to use more than two or three specimens when assessing measurement error on a subsample of the actual sample.

Conclusion

In this study, most artefactual variance in superimposed Landmark Dataset 1 (acquired from 'reasonable' surface variants) was caused by intra-observer error. The observed error between the days in Landmark Dataset 2 and the learning effect observed in Landmark Dataset 3 suggest that landmark data should be acquired by an experienced observer in as few sessions as possible; in this context, however, the effects of landmarking a high number of surfaces in the same session should be considered, as this might cause higher error due to fatigue. Intra-observer error can also be minimized by using surfaces that allow for a high precision in landmark placing. Results from Landmark Datasets 2 and 4 suggest that such surfaces can be obtained by careful surface simplification as long as the resolutions of the underlying volumes are good enough, and possibly also by applying automatic local thresholding with an optimal algorithm-parameter combination.

The second largest amount of artefactual variance in superimposed Landmark Dataset 1 was due to inter-observer error. Thus, it would be preferable that all landmark sets were acquired by only one experienced observer to avoid this type of measurement error. If landmark sets need to be acquired by multiple observers, it is important to assure that all

observers are well trained and that they are placing the landmarks consistently at homologous points.

In the superimposed Landmark Dataset 1, the contribution of the segmentation strategy for surface generation to the total variance has been significant yet small compared with other factors. Still, the strong effect of the factor segmentation in Landmark Dataset 2 implies that the segmentation strategy should be carefully chosen to obtain a surface that represents the natural shape of the specimen with best morphological fidelity. Although downsampling had no significant effect in this study, there are indicators that using a reasonable high spatial resolution for CT scanning is advisable for best morphological fidelity and highest landmark precision.

Despite the significant amount of measurement error in landmark data acquired from our 'reasonable' surface variants, the artefactual variation was still small relative to true biological shape differences. The observed 6.75% of artefactual variance probably could have been reduced by following the recommendations above. In our opinion, this small amount of measurement error and the potential for its further reduction justify the use of µCT-derived surfaces for 3D landmark data acquisition of anuran pectoral girdles for shape analysis between specimens and of variation between body halves within a given specimen.

Our experimental design revealed several options to reduce measurement error in the analysis of the anuran pectoral girdle shape by means of µCT-based geometric morphometrics. These options may well apply to other biological objects; however, we still follow previous recommendations (e.g. Klingenberg, 2015; Fruciano, 2016; Robinson & Terhune, 2017) to assess systematically the effects of all factors potentially contributing to measurement error in landmark data and to compare the magnitude of artefactual variance to the biological variation of interest.

Acknowledgements

We sincerely thank Juliana Lutz, Lena Schwinger, and Mehria Sedik for participating in landmark data acquisition, and Frank Friedrich for fruitful discussions. We are grateful to YXLON International GmbH and Helmholtz-Zentrum Geesthacht for granting access to their respective µCT systems. This work was funded by the Deutsche Forschungsgemeinschaft (HA 2323/14-1) and the Wilhelm-Peters-Fonds of the Deutsche Gesellschaft für Herpetologie und Terrarienkunde e.V.

Disclosure of interests

The authors declare no conflict of interests.

Author contributions

KE, JH, and AH designed the study. AB, SB, JUH, KE, and TK were involved in µCT scanning; SG provided infrastructure and additional expertise for µCT. Landmark sets were acquired by KE, JH, JUH, and, not appearing as authors, by Juliana Lutz, Lena Schwinger, and Mehria Sedik. KE generated the surfaces, performed the analyses, and drafted the manuscript. AH and JH commented on early versions of the manuscript. All authors critically revised the manuscript and approved the final version.

References

Adams DC, Collyer ML, Kaliontzopoulou A, Sherratt E (2017) *Geomorph: software for geometric morphometric analyses*. Available from: https://cran.r-project.org/package=geomorph.

Adams DC, Rohlf FJ, Slice DE (2013) A field comes of age: geometric morphometrics in the 21st century. *Hystrix* **24**, 7–14.

Adams DC, Rohlf FJ, Slice DE (2004) Geometric morphometrics: ten years of progress following the 'revolution'. *Ital J Zool* **71**, 5–16.

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26**, 32–46.

Anderson MJ, ter Braak, Cajo J. F. (2003) Permutation tests for multi-factorial analysis of variance. *J Stat Comput Sim* **73**, 85–113.

Arnqvist G, Mårtensson T (1998) Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. *Acta Zool Acad Sci Hung* **44**, 73–96.

Barbeito-Andrés J, Anzelmo M, Ventrice F, Sardi ML (2012) Measurement error of 3D cranial landmarks of an ontogenetic sample using computed tomography. *J Oral Biol Craniofac Res* **2**, 77–82.

Barbeito-Andrés J, Bernal V, Gonzalez PN (2016) Morphological asymmetries of mouse brain assessed by geometric morphometric analysis of MRI data. *Magn Reson Imaging* **34**, 980–989.

Bilfeld MF, Dedouit F, Sans N, Rousseau H, Rougé D, Telmon N (2013) Ontogeny of size and shape sexual dimorphism in the ilium: a multislice computed tomography study by geometric morphometry. *J Forensic Sci* **58**, 303–310.

Bonneau N, Bouhallier J, Simonis C, Baylac M, Gagey O, Tardieu C (2012) Technical note: shape variability induced by reassembly of human pelvic bones. *Am J Phys Anthropol* **148**, 139–147.

Bookstein FL (2003) *Morphometric tools for landmark data: geometry and biology*. Cambridge: Cambridge University Press.

Bookstein FL (1997) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med Image Anal* **1**, 225–243.

Buser TJ, Sidlauskas BL, Summers AP (2018) 2D or not 2D? Testing the utility of 2D vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). *Anat Rec* **301**, 806–818.

Cardini A (2014) Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures? *Hystrix* **25**, 73–81.

Chang J, Alfaro ME (2016) Crowdsourced geometric morphometrics enable rapid large-scale collection and analysis of phenotypic data. *Methods Ecol Evol* **7**, 472–482.

Christiansen BA (2016) Effect of micro-computed tomography voxel size and segmentation method on trabecular bone microstructure measures in mice. *Bone Rep* **5**, 136–140.

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G (2008) MeshLab: an open-source mesh processing tool. Sixth Eurographics Italian Chapter Conference, 129–136.

Coleman MN, Colbert MW (2007) Technical note: CT thresholding protocols for taking measurements on three-dimensional models. *Am J Phys Anthropol* **133**, 723–725.

Collard M, O'Higgins P (2001) Ontogeny and homoplasy in the papionin monkey face. *Evol Dev* **3**, 322–331.

Colman KL, Dobbe JGG, Stull KE, et al. (2017) The geometrical precision of virtual bone models derived from clinical computed tomography data for forensic anthropology. *Int J Legal Med* **131**, 1155–1163.

Corner BD, Lele S, Richtsmeier JT (1992) Measuring precision of three-dimensional landmark data. *J Quant Anthropol* **3**, 347–359.

Cox PG, Fagan MJ, Rayfield EJ, Jeffery N (2011) Finite element modelling of squirrel, guinea pig and rat skulls: using geometric morphometrics to assess sensitivity. *J Anat* **219**, 696–709.

Curth S, Fischer MS, Kupczik K (2017) Can skull form predict the shape of the temporomandibular joint? A study using geometric morphometrics on the skulls of wolves and domestic dogs. Ann Anat 214, 53–62.

Dryden IL (2017) *Shapes: statistical shape analysis*. Available from: https://CRAN.R-project.org/package=shapes.

Dryden IL, Mardia KV (2016) *Statistical shape analysis with applications in R. Wiley series in probability and statistics*. Chichester, UK, Hoboken, NJ: John Wiley & Sons.

Daboul A, Ivanovska T, Bülow R, Biffar R, Cardini A (2018) Procrustes-based geometric morphometrics on MRI images: An example of inter-operator bias in 3D landmarks and its impact on big datasets. PLoS ONE 13, e0197675.

Engelkes K, Friedrich F, Hammel JU, Haas A (2018) A simple setup for episcopic microtomy and a digital image processing workflow to acquire high-quality volume data and 3D surface models of small vertebrates. *Zoomorphology* **137**, 213–228.

Fajardo RJ, Ryan T, Kappelman J (2002) Assessing the accuracy of high-resolution x-ray computed tomography of primate trabecular bone by comparisons with histological sections. *Am J Phys Anthropol* **118**, 1–10.

Fruciano C (2016) Measurement error in geometric morphometrics. *Dev Genes Evol* **226**, 139–158.

Fruciano C, Celik MA, Butler K, Dooley T, Weisbecker V, Phillips MJ (2017) Sharing is caring? Measurement error and the issues arising from combining 3D morphometric datasets. Ecol Evol 7, 7034–7046.

Gunz P, Mitteroecker P, Bookstein FL (2005) Semilandmarks in three dimensions. In: *Modern morphometrics in physical anthropology. Developments in primatology* (ed Slice DE), pp. 73–98. New York: Kluwer Academic/Plenum Publishers.

Gunz P, Ramsier M, Kuhrig M, Hublin J, Spoor F (2012) The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. *J Anat* **220**, 529–543.

Hale AR, Honeycutt KK, Ross AH (2014) A geometric morphometric validation study of computed tomography–extracted craniofacial landmarks. *J Craniofac Surg* **25**, 231–237.

Healy S, McMahon J, Owens P, Dockery P, FitzGerald U (2018) Threshold-based segmentation of fluorescent and chromogenic images of microglia, astrocytes and oligodendrocytes in FIJI. *J Neurosci Meth* **295**, 87–103.

Heuzé Y, Kawasaki K, Schwarz T, Schoenebeck JJ, Richtsmeier JT (2016) Developmental and evolutionary significance of the zygomatic bone. *Anat Rec* **299**, 1616–1630.

Hildebolt CF, Vannier MW, Knapp RH (1990) Validation study of skull three-dimensional computerized tomography measurements. *Am J Phys Anthropol* **82**, 283–294.

Kassambara A (2017) *Gpubr: 'ggplot2' based publication ready plots*. Available from: https://CRAN.R-project.org/package=ggpubr.

Kesterke MJ, Judd MA, Mooney MP, et al. (2018) Maternal environment and craniofacial growth: geometric morphometric analysis of mandibular shape changes with in utero thyroxine overexposure in mice. *J Anat* **233**, 46–54.

Klingenberg C (2015) Analyzing fluctuating asymmetry with geometric morphometrics: concepts, methods, and applications. *Symmetry* **7**, 843–934.

Klingenberg CP, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* **56**, 1909–1920.

Kulemeyer C, Asbahr K, Gunz P, Frahnert S, Bairlein F (2009) Functional morphology and integration of corvid skulls - a 3D geometric morphometric approach. *Front Zool* **6**, 2.

Landini G, Randell DA, Fouad S, Galton A (2017) Automatic thresholding from the gradients of region boundaries. *J Microsc* **265**, 185–195.

Lee JC (1982) Accuracy and precision in anuran morphometrics: artifacts of preservation. *Syst Zool* **31**, 266–281.

Lemon J (2006) Plotrix: a package in the red light district of R. *R-News* **6**, 8–12.

Lockwood CA, Lynch JM, Kimbel WH (2002) Quantifying temporal bone morphology of great apes and humans: an approach using geometric morphometrics. *J Anat* **201**, 447–464.

Lorkiewicz-Muszyńska D, Kociemba W, Sroka A, et al. (2015) Accuracy of the anthropometric measurements of skeletonized skulls with corresponding measurements of their 3D reconstructions obtained by CT scanning. *Anthropol Anz* **72**, 293–301.

Maglia AM, Púgener LA (1998) Skeletal development and adult osteology of *Bombina orientalis* (Anura: Bombinatoridae). *Herpetologica* **54**, 344–363.

Marcy AE, Fruciano C, Phillips MJ, Mardon K, Weisbecker V (2018) Low resolution scans can provide a sufficiently accurate, cost- and time-effective alternative to high resolution scans for 3D shape analyses. *PeerJ* 6, e5032.

McGill R, Tukey JW, Larsen WA (1978) Variations of box plots. *Am Stat* **32**, 12–16.

Mitteroecker P, Gunz P, Windhager S, Schaefer K (2013) A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix* **24**, 59–66.

Monteiro L (2000) Geometric morphometrics and the development of complex structures: ontogenetic changes in scapular shape of dasypodid armadillos. *Hystrix* **11**, 91–98.

O'Higgins P, Jones N (1998) Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. *J Anat* **193**, 251–272.

Oksanen J, Blanchet FG, Friendly M, et al. (2017) *Vegan: community ecology package*. Available from: https://CRAN.R-project.org/package=vegan.

Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9**, 62–66.

Parkinson IH, Badiei A, Fazzalari NL (2008) Variation in segmentation of bone from micro-CT imaging: implications for quantitative morphometric analysis. *Australas Phys Eng Sci Med* **31**, 160–164.

Perez SI, Bernal V, Gonzalez PN (2006) Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *J Anat* **208**, 769–784.

Plate T, Heiberger R (2016) *Abind: combine multidimensional arrays*. Available from: https://CRAN.R-project.org/package=abind.

Pujol A, Rissech C, Ventura J, Badosa J, Turbón D (2014) Ontogeny of the female femur: geometric morphometric analysis applied on current living individuals of a Spanish population. *J Anat* **225**, 346–357.

R Core Team (2017) *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Rabinovich SG (2006) *Measurement Errors and Uncertainties: Theory and Practice*. New York, NY: Springer Science and Media, Inc.

Richtsmeier JT, Paik CH, Elfert PC, Cole TM, Dahlman HR (1995) Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans. *Cleft Palate-Cran J* **32**, 217–227.

Robinson C, Terhune CE (2017) Error in geometric morphometric data collection: combining data from multiple sources. *Am J Phys Anthropol* **164**, 62–75.

Ross AH, Williams S (2008) Testing repeatability and error of coordinate landmark data acquired from crania. *J Forensic Sci* **53**, 782–785.

RStudio Team (2017) *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. Available from: http://www.rstudio.com/.

Schindelin J, Arganda-Carreras I, Frise E, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging* **13**, 146–165.

Shearer BM, Cooke SB, Halenar LB, et al. (2017) Evaluating causes of error in landmark-based data collection using scanners. *PLoS ONE* **12**, e0187452.

Simon MN, Marroig G (2015) Landmark precision and reliability and accuracy of linear distances estimated by using 3D computed micro-tomography and the open-source TINA Manual Landmarking Tool software. *Front Zool* **12**, 12.

Singleton M (2002) Patterns of cranial shape variation in the Papionini (Primates: Cercopithecinae). *J Hum Evol* **42**, 547–578.

Sokal RR, Rohlf FJ (1981) *Biometry: The principles and practice of statistics in biological research*. San Francisco: Freeman.

Valeri CJ, Cole TM, Lele S, Richtsmeier JT (1998) Capturing data from three-dimensional surfaces using fuzzy landmarks. *Am J Phys Anthropol* **107**, 113–124.

Verhaegen Y, Adriaens D, Wolf TD, Dhert P, Sorgeloos P (2007) Deformities in larval gilthead sea bream (*Sparus aurata*): a qualitative and quantitative analysis using geometric morphometrics. *Aquaculture* **268**, 156–168.

von Cramon-Taubadel N, Frazier BC, Lahr MM (2007) The problem of assessing landmark error in geometric morphometrics: theory, methods, and modifications. *Am J Phys Anthropol* **134**, 24–35.

Wang C, Hsu H, Wang C, Lee T, Kuo Y (2015) Quantifying floral shape variation in 3D using microcomputed tomography: a case study of a hybrid line between actinomorphic and zygomorphic flowers. *Front Plant Sci* **6**, 724.

Wickham H (2016) *Ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag. Available from: http://ggplot2.org.

Wiley DF, Amenta N, Alcantara DA, et al. (2005) Evolutionary morphing. *IEEE Visualization*.

Williams FL, Richtsmeier JT (2003) Comparison of mandibular landmarks from computed tomography and 3D digitizer data. *Clin Anat* **16**, 494–500.

Wilson LAB, Cardoso HFV, Humphrey LT (2011) On the reliability of a geometric morphometric approach to sex determination: a blind test of six criteria of the juvenile ilium. Forensic Science International 206, 35–42.

Zelditch M, Swiderski DL, Sheets HD (2012) *Geometric morphometrics for biologists: a primer*. Amsterdam, Boston: Elsevier Academic Press.

Supporting Information

[The following Supporting Information can be found at the end of this thesis, pp. 233-243.]

Fig. S1. Bones of *Bombina orientalis* (ZMH A12601).

Fig. S2. Steps to determine the optimal segmentation strategy for the µCT scan of a selected *Bombina orientalis* specimen (ZMH A12601) using automatic local thresholding.

Table S1. Thresholding quality assessed for bone volumes and the adjacent two rows of voxels (first quality measure) of automatic local thresholding trials.

Table S2. Landmark definitions and visualizations.

Table S3. Overview of the composition of Landmark Datasets 1–4.

Text S1. Explanations on workflow to generate a synthetic CT volume (‘phantom stack’) and to determine the case-specific best combination of automatic local thresholding algorithm and parameters.

- Chapter three -

Accuracy of bone segmentation and surface generation strategies analyzed by using synthetic CT volumes

Karolin Engelkes[1]

1) Center of Natural History (CeNak), Universität Hamburg, Hamburg, Germany

This manuscript is publicly available as a preprint (doi 10.13140/RG.2.2.18389.86247) and has been submitted to the Journal of Anatomy.

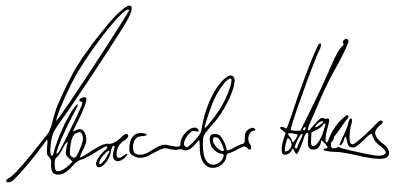
Contributions of KE:

Study design: 100%

Data collection: 100%

Data analysis: 100%

Manuscript drafting: 100%

Contributions confirmed by supervisor: 

Prof. Dr. Alexander Haas

Abstract

Different kinds of bone measurements are commonly derived from computed-tomography (CT) volumes to answer a multitude of questions in biology and related fields. The underlying steps of bone segmentation and, optionally, polygon surface generation are crucial to keep the measurement error small. In this study, the performance of different, easily accessible segmentation techniques (global thresholding, automatic local thresholding, weighted random walk, neural network, watershed) and surface generation approaches (different algorithms combined with varying degrees of simplification) was analyzed and recommendations for minimizing inaccuracies were derived. The different approaches were applied to synthetic CT volumes for which the correct segmentation and surface geometry was known. The most accurate segmentations of the synthetic volumes were achieved by setting a case-specific window to the gray value histogram and subsequently applying automatic local thresholding with appropriately chosen thresholding method and radius. Surfaces generated by the Amira® module *Generate Lego Surface* in combination with careful surface simplification were the most accurate. Surfaces with sub-voxel accuracy were obtained even for synthetic CT volumes with low contrast to noise ratios. Segmentation trials with real CT volumes supported the findings. Very accurate segmentations and surfaces can be derived from CT volumes by using readily accessible software packages. The presented results and derived recommendations will help to reduce the measurement error in future studies. Further, the demonstrated strategies for assessing segmentation and surface qualities can be adopted to quantify the performance of new segmentation approaches in future studies.

Keywords: 3D reconstruction, accessible segmentation techniques, local thresholding, segmentation quality, sub-voxel accuracy, surface accuracy, surface simplification

Introduction

Different kinds of bone measurements like coordinates of landmarks, linear distances, or variables describing trabecular bone are routinely used in biological studies and related fields to answer a multitude of questions. Such measurements are acquired from, among others, computed tomography (CT) volumes or derived polygon surfaces (e.g., Spoor et al. 1993; Bouxsein et al. 2010; Cornette et al. 2013; Andjelković et al. 2016). Prior to performing the actual measuring, there usually is a step of segmenting the bones of interest and, optionally, a step of polygon surfaces generation (including surface simplification, i.e., reduction of polygon count and smoothing). Various studies (e.g., Fajardo et al. 2002; Barandiaran et al. 2009; Rathnayaka et al. 2011; Christiansen, 2016; Engelkes et al. 2019; Ito, 2019) have shown that the choices of segmentation and surface generation strategies have considerable or even statistically significant effects on the measured values.

In CT volumes there is a gradual change of gray values (i.e., x-ray attenuation coefficients) at the boundary of adjacent structures with differing x-ray attenuation coefficients (Koehler et al. 1979; Spoor et al. 1993; Coleman & Colbert, 2007; Supporting

Information Fig. S1). This gray value gradient is caused by optical effects during x-ray imaging (Witt et al. 2003), the back projection of the x-ray images for volume reconstruction (Buzug, 2008), and by partial volume averaging (i.e., the gray value of each voxel is the average of the tissue densities within it; Fajardo et al. 2002; Barrett & Keat, 2004; Abel et al. 2013). The resulting gradient between anatomical structures makes it challenging to determine the correct position of the boundaries. Previous studies (Ullrich et al. 1980; Baxter & Sorenson, 1981; Magnusson, 1987) have shown that the correct boundary position can be calculated from the gray values of the respective structures; the correct boundary is located at the mean value (called 'half maximum height', 'HMH') of the minimum and maximum gray values along a line that crosses the boundary (Spoor et al. 1993; Prevrhal et al. 1999; Coleman & Colbert, 2007).

In CT volumes, the gray values may differ within and between bones. Such differences can result from a natural inhomogeneity in the mineral density (Lindh et al. 2004; Roschger et al. 2008), from the volume being blurred by the point-spread function (PSF; i.e., structures that are thin relative to the width of the PSF are represented by darker gray values; Prevrhal et al. 1999), from partial volume averaging (i.e., darker gray values in regions of thin bone; Fajardo et al. 2002; Roschger et al. 2008), or from other artifacts that produce locally darker or brighter gray values (see, e.g., Barrett & Keat, 2004; Boas & Fleischmann, 2012; Keklikoglou et al. 2019). In addition, the basic intensity of the gray values might vary within a CT volume (e.g., cupping artifact; Barrett & Keat, 2004).

Various approaches for CT volume segmentation have been proposed (e.g., Baillard & Barillot, 2000; Lamecker et al. 2004; Gelaude et al. 2006; Scherf & Tilgner, 2009; Minnema et al. 2018). Yet, in the field of biology, global thresholding seems to be the most widely used approach to separate bone voxels from surrounding tissues; a global threshold is either subjectively selected by eye (e.g., Cornette et al. 2013; Andjelković et al. 2016) or computed based on the gray value distribution of the volume (e.g., Fajardo et al. 2002; Guyomarc'h et al. 2012). Given that the gray values of bones vary locally, the optimal threshold for obtaining an accurate segmentation likely differs within and between bones (Fajardo et al. 2002; Rathnayaka et al. 2011) and global thresholding presumably does not result in the best achievable segmentation quality. This is supported by a recent study (Ito, 2019) in which a gradient-based watershed algorithm outperformed global thresholding with regard to the segmentation quality.

Once bone voxels have been segmented, polygon surfaces can be generated. Various software packages with different surface generation algorithms are available to accomplish

this task. In addition, choices on surface simplification have to be made. Simplification has been shown to have case-specific positive or negative effects on the surface accuracy (e.g., Bade et al. 2006; DeVries et al. 2008; Veneziano et al. 2018), while the effects of the surface generation algorithms are not well understood.

In this study, synthetic CT volumes of varying image qualities were created; the correct segmentation and surface geometry was known for these volumes. The performance of different segmentation and surface generation strategies was assessed to determine the approach that resulted in the best quality. The approaches compared herein were selected because of their accessibility in freely available or widely used commercial software and their ease of use. The observations were verified for real CT volumes. The results and recommendations will help to minimize the measurement error in future studies in which measurements of bones are acquired from CT volumes or derived surfaces.

Materials and methods

Generation of synthetic CT volumes

A preexisting CT volume (Engelkes et al. 2019) of a *Bombina orientalis* (Boulenger, 1890) specimen (ZMH A12601; voxel size arbitrarily set to 1) was chosen to generate surfaces of the pectoral girdle bones. The bones were segmented in Amira® (version 6.0.1; Konrad-Zuse-Zentrum Berlin, FEI Visualization Sciences Group; *Magic Wand* tool, correction with *Brush* tool) and simplified polygon surfaces were exported with the *MultiExport* macro that combines the Amira®-functions *Isosurface*, *Extract Surface*, *Simplification Editor*, and *Smooth Surface* (see Engelkes et al. 2018 for details on the macro). The surfaces were cleaned and combined into one surface object in MeshLab (version 2016.12; used filters: *Surface Reconstruction: VCG*, *Taubin Smooth*, *Simplification: Quadric Edge Collapse Decimation*; Cignoni et al. 2008). The finally processed surface that contained all pectoral girdle bones will subsequently be called ‘phantom surface’ (Supporting Information Fig. S2A).

The phantom surface was converted to a gray scale volume of 600x600x312 voxels (Amira®, *Scan Surface To Volume*; resolution corresponded to original CT volume, voxel size arbitrarily set to 1) and areas mimicking parts of other bones and soft tissue were added (*Arithmetic*). The resulting volume (‘phantom stack/volume’) contained bright bone, medium gray soft tissue and dark air voxels with sharp and, thus, well-known boundaries between structures (Supporting Information Fig. S2B).

Cone beam CT scans and back projections of the phantom stack were mathematically simulated using the ASTRA toolbox (version 1.8; van Aarle et al. 2015; van Aarle et al. 2016) in Matlab® (version R2018b; The MathWorks; Supporting Information Fig. S3); sinograms were generated (*astra_create_sino3d_cuda* function) and volumes were reconstructed using either the *FDK_CUDA* ('FDK') or the *SIRT3D_CUDA* ('SIRT') algorithm (*astra_mex_algorithm* function). The image noise of an arbitrarily chosen preexisting CT volume was extracted and added to each of the two synthetic volumes in Fiji (based on ImageJ 1.51n; Schindelin et al. 2012; Schneider et al. 2012); various volume variants with varying noise intensities were created. Additionally, various levels of Poisson noise were added to the sinograms (Matlab®, *astra_add_noise_to_sino* function) and volumes were reconstructed as above. This resulted in 14 synthetic CT volumes (16 bit; published as Engelkes, 2020; downloadable from http://doi.org/10.25592/uhhfdm.962) with either a short (FDK) or a long (SIRT) gray gradient between structures, with rather homogeneous (FDK) or heterogeneous (SIRT) gray value distributions within bones, and with differing noise characteristics and intensities (Supporting Information Fig. S4). It should be noted that the tissue gray values in the phantom stack were arbitrarily chosen; therefore the gray values in the synthetic CT volumes have no physical meaning (e.g., in terms of Hounsfield units).

The mean gray values of the pectoral girdle bones and the soft tissue, as well as the standard deviation of the soft tissue gray values, were assessed for each synthetic CT volume in Amira® (*Material Statistics* module). Soft tissue voxels were arbitrarily segmented for evaluation. The voxels of the pectoral girdle bones were segmented using a low threshold; the mean gray value of these voxels was determined, then the segmented region was shrunk (i.e., eroded in terms of mathematical morphology) by one layer of voxels (*shrink* tool) and the new mean gray value was assessed. This process was repeated until the maximal mean gray value of bone was obtained. This maximal value was used to calculate the contrast to noise ratio (CNR) of bone and soft tissue by dividing the difference of the mean gray values of bone and soft tissue by the standard deviation of the soft tissue.

Thresholding-based segmentation

The pectoral girdle bones were segmented in each synthetic CT volume using global and local thresholding (Supporting Information Fig. S5). Various trials of global thresholding were performed for each volume with thresholds varying by 50 gray scale units between trials (*Magic Wand* tool in Amira®). Thresholding trials were repeated until the best possible segmentation quality was achieved for each synthetic volume. The quality of each

segmentation result was measured as the percentage of correctly classified voxels within the inner and outer two voxel layers adjacent to the correct boundary of bones that was inferred from the phantom stack (Supporting Information Fig. S2C). This measure of segmentation quality was chosen as it directly measures the quantity of interest and, in the context of this study, has no disadvantage compared to previously suggested measures (reviewed in Taha & Hanbury, 2015).

The segmentation quality achieved by automatic local thresholding was assessed, too. Local thresholding differs from global thresholding in that a threshold is determined for each pixel separately based on the values of the surrounding pixels in a neighborhood of a user-defined size. The local threshold can be calculated by methods previously suggested for global thresholding (e.g., Otsu, 1979; Phansalkar et al. 2011). Before the actual local thresholding, different versions of each synthetic CT volume were generated by systematically varying the lower and upper value at which the stack histogram was cut off (Fiji). The centers set for these histogram windows were changed by 50, the window widths by 100 gray scale units between trials. The adjusted volumes were converted to 8 bit and all (except for *Sauvola* which did not result in reasonable binarizations in an informal pre-test) local thresholding methods implemented in the Fiji plugin *Auto Local Threshold* (Landini, Rueden, Schindelin, Hiner, & Helfrich, https://imagej.net/Auto_Local_Threshold) were applied to the different versions of the volumes. The radius that determined the size of the local neighborhood was varied by one pixel between trials while all other parameters were kept at the default values. Thresholding was performed slice-wise using images showing cross sections of the specimen ('cross' approach).

The robustness of the automatic local thresholding methods to deviations from the optimal values of either the histogram window center or widths, or the radius was assessed. Any one of these parameters was altered by either 100 gray scale units (center, width) or one (radius), while the others were set to their optimum. The difference between the thresholding quality achieved with the non-optimal parameter and the quality achieved with the optimal parameters was determined. The mean decrease of the segmentation quality across synthetic CT volumes was used as indicator for the robustness of the respective automatic local thresholding method. Linear models were fitted (*lm* function in R version 3.6.0 via RStudio version 1.1.463; R Core Team, 2019; RStudio Team, 2018) to assess the dependence of the optimal window center on the mean gray value of bone and soft tissue for each thresholding method.

Selected synthetic stacks were resliced (*Reslice [/]…* function in Fiji) from top to bottom and from left to right. Automatic local thresholding with the method *Otsu* and the parameter combination that was optimal for the respective volume was applied to each resliced stack. All three thresholding results of a given volume were combined using two different voting rules for decision fusion (e.g., Mitchell, 2010) to derive two segmentations: a given voxel was classified as bone if it was recognized as bone in any two out of the three ('2of3'; majority voting sensu Mitchell, 2010) or in all three ('3of3') thresholded volumes.

Watershed segmentation

The watershed segmentation approach uses an image gradient (i.e., the local change of gray values in the CT volume) and results in a segmentation in which the boundaries of adjacent structures are set at the position of the highest gradient (Vincent & Soille, 1991). This approach was used to segment selected synthetic CT volumes with the *Watershed Segmentation* function of Amira®. Several image gradients were calculated for each selected stack using the Fiji-functions *Gradient (3D)* or *Canny Edge* (detection mode *Volumetric*, other parameters as default; XLib, https://imagej.net/Xlib), the *Sobel Filter* (applied in 3D, Amira®), or generated directly in the *Watershed Segmentation* function.

Seed regions for the background and soft tissue were generated by shrinking the respective regions of the phantom stack by two layers of voxels. Three different seeds were used for the pectoral girdle bones: bone voxels in the best segmentation result of automatic local thresholding (*Otsu* method) shrunk by one layer of voxels ('Otsu-1-Seeds', calculated for each volume separately), and the respective region of the phantom stack shrunk by one or two layers of voxels ('Phantom-1-Seeds', 'Phantom-2-Seeds', respectively).

Segmentation using Biomedisa

The two segmentation approaches implemented in the Biomedical Image Segmentation App (Biomedisa; Lösel & Heuveline, 2016) were tested for selected synthetic CT volumes. An image stack with four unsegmented slices between segmentations of the phantom stack was used for the weighted random walk approach (segmentation performed on several days in November 2019). A large number of weighted random walks that started from the labeled slices was performed by Biomedisa; each voxel in the volume was hit by several random walks and a given voxel was assigned to the label from which the most random walks hitting the voxel started (Lösel & Heuveline, 2016).

In the second segmentation approach in Biomedisa, a neural network was trained and refined (all parameters set to their default values; training and segmentation performed on several days in May 2020). The synthetic volumes that contained no and highest (among selected volumes) image noise of each reconstruction algorithm (four volumes in total) and an error-free segmentation of the phantom stack were used for the training. The selected synthetic CT volumes were segmented with the refined network.

Surface generation and accuracy

Several modules to generate polygon surfaces are implemented in Amira®: (i) *Generate Lego Surface*, (ii) *Isosurface* in combination with *Extract Surface*, and (iii) *Generate Surface*. The latter allows surface simplification, while the former two approaches do not. All surfaces generated by the former two approaches were simplified by iteratively reducing the polygon count (*Simplification Editor*) to about 90 % (per iteration) and smoothing the surface (*Smooth Surface*).

In a first step, the surface generation approach that resulted in the most accurate surface was assessed. The phantom surface was used to generate three volumes that differed in their voxel size (Amira®). Voxel sizes of 0.5, 1, and 2 units were used; the voxel size of 1 unit corresponded to the resolution of the original CT volume. For each volume, several surface variants of the pectoral girdle bones were generated by approaches (i) through (iii); surface simplification ranged from no simplification to strong simplification. For the third approach (iii), the effects of *Constrained* and *Unconstrained Smoothing*, and compactification (i.e., option *Compactify* either checked or unchecked) were compared.

The accuracies of the derived surface were assessed by calculating the distance of each vertex of the phantom surface (ground truth) to the closest vertex of the generated surfaces (Amira®, *Surface Distance*). Prior to distance calculation, the polygons of the derived surfaces were subdivided (MODO® version 10.1v2; The Foundry; *Subdivide: Faceted* function) in order to increase the vertex count.

The surface generation approach that resulted in the most accurate surface was used to generate (simplified) surfaces of selected (i.e., best) segmentation results of the synthetic CT volumes. The accuracies of the derived surfaces were assessed as above.

Application to real CT volumes

Three µCT volumes of frogs (Supporting Information Table S1) were downloaded from MorphoSource (Duke University, https://www.morphosource.org/) to test the performance of

the segmentation strategies compared herein. Voxel sizes were arbitrarily set to one and the volumes were cropped such that they contained a similar anatomical region as the phantom stack. The segmentation approaches selected above were applied to the sub-volumes and the segmentation qualities were indirectly measured by estimating the accuracy of derived polygon surfaces.

In particular, the parameters characterizing the image quality were determined as above and used to approximate the optimal parameters for automatic local thresholding based on the results illustrated in Fig. 1. As estimating the window width from Fig. 1D resulted in many bone voxels disappearing by getting black, a window width was subjectively determined by eye; a width was chosen that was as small as possible but wide enough such that no bone voxels disappeared. It should be noted that the window center possibly was sub-optimal for trabecular bone, because the voxels of trabeculae were not included in the volume finally considered for the calculation of the mean gray value of bone. The gray-value-adjusted µCT volumes and resliced versions of them were thresholded by automatic local thresholding with the method *Otsu* and a radius estimated from the corresponding CNR (compare Fig. 1D). Two segmentations were generated with the approaches “cross” and “2of3”.

Global thresholds varying by 500 gray scale units were used to segment the real CT volumes (Amira®). For each volume, the threshold that resulted in the highest estimated surface accuracy (see below) was used in the comparison of the segmentation accuracy among approaches.

For watershed segmentation, bone and soft tissue voxels were segmented by visually determined thresholds; the segmented regions were shrunk by one layer of voxels and used as seed regions (Amira®). Image gradients were calculated with the *Gradient (3D)* function (Fiji) and the *Sobel Filter* (applied in 3D, Amira®).

Label datasets that contained four unsegmented slices alternating with the segmentations of the pectoral girdle bones as obtained from automatic local thresholding (2of3 approach) were generated. Those labels were used to segment the real CT volumes with the weighted random walk approach implemented in Biomedisa. Additionally, the volumes were segmented with the neuronal network trained with the synthetic CT volumes (Biomedisa trials were performed on several days in May 2020).

The pectoral girdle bones were separated from calcified cartilage by a rough manual segmentation (Amira®). For each µCT volume, final segmentations of the pectoral girdle bones were created by combining the rough manual segmentation with the results of the different segmentation strategies.

Unsimplified surfaces were created with the *Generate Lego Surface* module (Amira®) and used to visually assess differences in the two segmentations obtained by automatic local thresholding (cross and 2of3). Simplified polygon surfaces were generated from all segmentations by using the *Generate Lego Surface* module in combination with surface simplification to a subjectively optimal degree (approach i). The accuracy of each surface was estimated by calculating the distance of 10.000 randomly selected vertices to the positions of the local HMH along the corresponding surface normal (measure inside by the measure used in Ito, 2019). Therefore, the gray values along a given normal were extracted (distance between points of extraction: 0.1 voxel; maximum distance to vertex: 3 voxels; Amira®) and interpolated with a B-spline (degree = 5; *bs* function of *splines* package in R). The position of the HMH was determined as the point at which the value of the B-spline function equaled the HMH value (determined to the nearest 0.01 voxel; Supporting Information Fig. S6). It should be noted that this surface accuracy estimation might be affected by errors in B-spline interpolation, inaccuracies in the determination of the HMH, and by image noise. The estimated accuracies of the surfaces were used to indirectly measure the qualities of the respectively underlying segmentations.

Data visualization

Plots were generated in R (basic functions, *ggplot2* version 3.1.1; Wickham, 2016). The per-vertex distances of selected surfaces from the phantom surface were visualized in GOM Inspect 2018 (GOM GmbH). Randomly selected vertices that deviated more than 0.75 voxel from the position of the HMH value in the real CT volumes were visualized in MODO. Figures were created in Adobe® Illustrator® CS6 (version 16.0.3; Adobe® Systems Software).

Results and discussion

Automatic local thresholding

If applied with case-specifically optimal parameters, all methods of automatic local thresholding resulted in comparable segmentation qualities for a given synthetic CT volume (Fig. 1A; Supporting Information Table S2). The achievable quality decreased with the CNR of bone and soft tissue. Thus, best thresholding results can be obtained for CT stacks with high contrasts and low image noise. The use of high quality CT volumes is therefore recommendable if measurements should be taken from the volumes or derived surfaces.

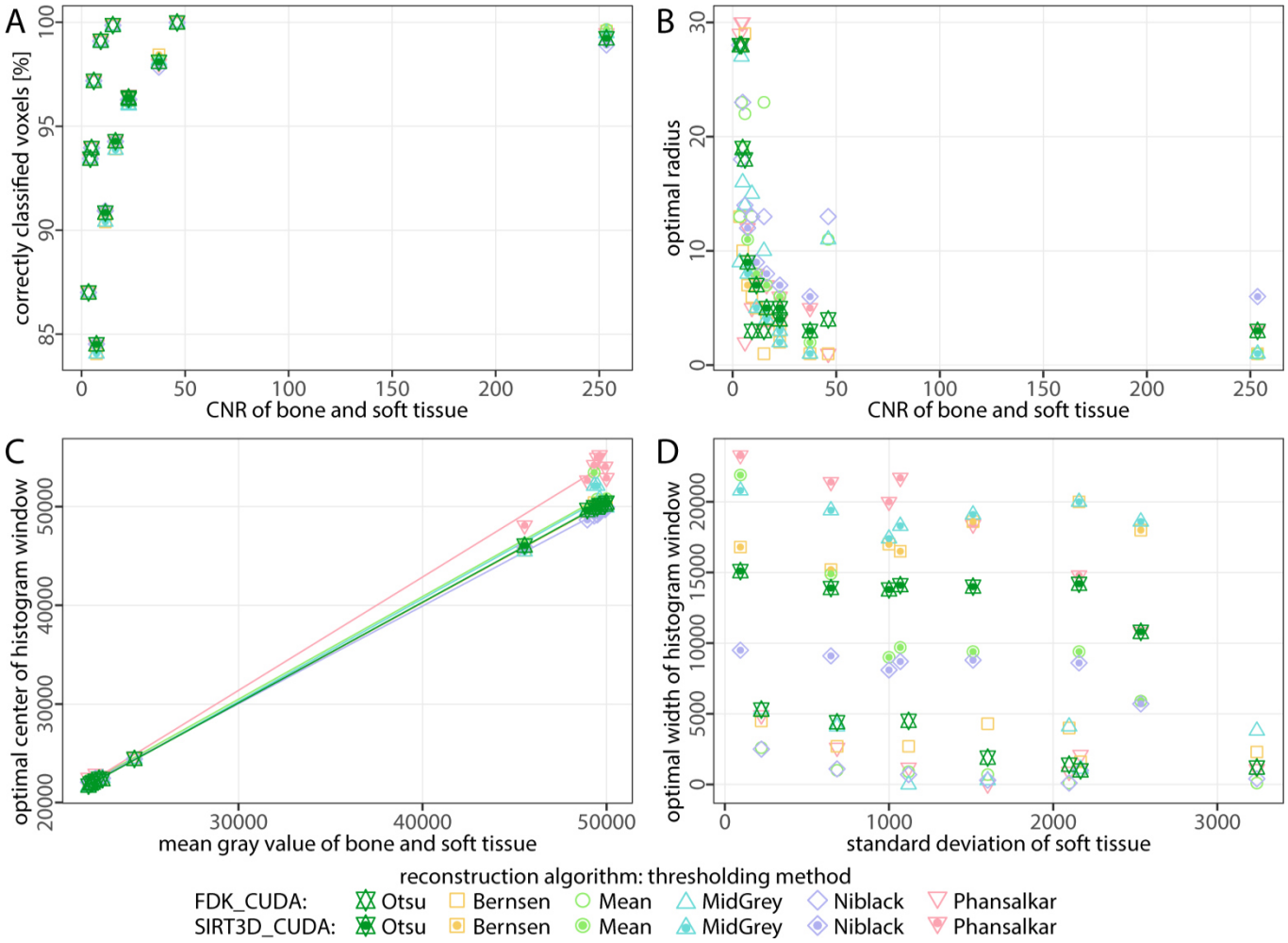


Fig. 1 Quality of automatic local thresholding and dependence of optimal thresholding parameters on characteristics of synthetic CT volumes (raw values can be found in Supporting Information Table S2). (A) Segmentation quality (measured as percentage of correctly classified voxels at the boundary of bone ± 2 layers of voxels) in dependence of the contrast to noise ratio (CNR) of bone and soft tissue. (B) Optimal radius for automatic local thresholding in dependence of the CNR. The approximately hyperbolic dependence can be used to estimate the optimal radius for any CT volume. (C) Optimal center of histogram window in dependence of the mean gray value of bone and soft tissue, as well as fitted lines. The linear dependence can be used to estimate the optimal window center for any CT volume. (D) Optimal width of histogram window in dependence of the standard deviation of soft tissue.

The quality of automatic local thresholding considerably depended on the histogram window that was set prior to thresholding; the window center had a higher impact on the results than the width (i.e., deviations from the case-specific optimal center resulted in a larger decrease in the thresholding quality than deviations from the optimal width; Table 1). These observations are in accordance with previous studies (Koehler et al. 1979; Seibert et al. 1981) that reported linear measurements by humans to be most accurate if the histogram window center for visualization was chosen appropriately, while the window width had a minor effect.

Table 1 Robustness of different automatic local thresholding methods to deviations from optimal parameters measured as mean decrease of segmentation quality (measured as percentage of correctly classified voxels at the boundary of bone ± 2 layers of voxels).

Method	Center ± 100[a] [%]	Width ± 100[b] [%]	radius ± 1[c] [%]
Otsu	-0.012763	-0.002958	-0.029248
Bernsen	-0.010874	-0.002137	-0.115276
Mean	-0.013682	-0.004868	-0.053011
MidGrey	-0.012493	-0.002557	-0.123375
Niblack	-0.013541	-0.004308	-0.046631
Phansalkar	-0.051889	-0.052353	-0.052116

[a] Optimal center of the histogram window varied by ±100 gray scale units
[b] Optimal width of the histogram window varied by ±100 gray scale units
[c] Optimal radius set for automatic local thresholding varied by ±1; thresholding quality for trials with radius of 2 for method *Otsu* excluded from calculations of mean decrease

The optimal center of the histogram window depended on the mean gray value of bone and soft tissue (Fig. 1C). The center could best be predicted for the method *Otsu* (fitted model: *center* = 1.019 * *mean* - 462.812; mean residual: 95.89; residual standard error: 126.7), followed by the method *Niblack* (fitted model: *center* = 0.9889 * *mean* + 374.50; mean residual: 125.34; residual standard error: 185). These regression functions can, thus, be used to calculate the optimal window center from the mean gray value of bone and soft tissue for any CT volume of interest. The optimal width of the histogram window (Fig. 1D) was just wide enough such that no bone voxel got black.

The optimal radius used for automatic local thresholding decreased non-linearly with an increase in the CNR (Fig. 1B). For the thresholding method *Otsu*, radii below three resulted in a disproportionately strong decrease in the thresholding quality; thus, a radius of three seems to be the lower limit for this method. Deviations from the case-specific optimal radius had a comparable high negative effect on the thresholding quality (Table 1). The decrease in the quality with deviation from the optimal radius was smallest for the method *Otsu*. Considering this and that the histogram window center was best predictable for the method *Otsu*, this method seems to be the best choice for automatic local thresholding. Previous studies (Landini et al. 2017; Healy et al. 2018), in which local versions of the thresholding algorithm by Ostu (1979) yielded good results for non-CT images, support the recommendation to use the method *Otsu* for automatic local thresholding of bones.

Segmentation quality across approaches

Automatic local thresholding with the method *Otsu* and optimal parameters resulted in a higher segmentation quality than the other segmentation approaches for most synthetic CT

Table 2 Volume-specific segmentation quality (measured as percentage of correctly classified voxels at the boundary of bone ± 2 layers of voxels) achieved with different segmentation approaches.

Characteristics of synthetic CT volumes					Segmentation quality [%]				
Reco[a]	Noise[b]	SD[c]	Mean[d]	CNR[e]	Local threshold[f]	Global threshold[g]	Water-shed[h]	Random walk[i]	Neural network[j]
FDK	No	222.1	21849	46.1	99.9928	99.9774	95.3783	98.4552	99.9665[k]
FDK	Real	1119.1	22072	9.3	99.1149	99.0842	95.2379	98.5776	98.7718
FDK	Real	2166.6	22235	4.8	93.9531	93.9484	94.5294	94.7434	93.6931
FDK	Poisson	2096.8	24342	4.2	93.4427	93.4397	94.1009	95.1581	97.5218[k]
SIRT	No	94.5	49325	253.5	99.2343	94.1470	94.1766	96.8264	99.8447[k]
SIRT	Poisson	999.7	48953	22.8	96.3169	92.4933	93.6455	94.9737	94.7104
SIRT	Real	1068.5	49621	22.7	96.3868	92.2796	93.6834	94.8764	95.9852
SIRT	Real	2157.4	49994	11.5	90.8477	88.3253	92.1341	90.5455	88.5019[k]

[a] Reconstruction algorithm used during generation of synthetic CT volumes
[b] Type of noise added to the synthetic CT volumes; No: no noise; Poisson: Poisson noise added to sinograms; Real: noise extracted from real CT volume added to synthetic stacks
[c] Standard deviation of soft tissue
[d] Mean gray value of bone and soft tissue
[e] Contrast to noise ratio of bone and soft tissue
[f] Segmentation by automatic local thresholding using method *Otsu* with volume-specific optimal parameter combination
[g] Segmentation by using the optimal global threshold
[h] Segmentation by watershed approach; gradient calculated by Fiji-function *Gradient (3D)* and Phantom-1-Seeds
[i] Segmentation by weighted random walk in Biomedisa; result type: regular
[j] Segmentation by neural network in Biomedisa; result type: regular
[k] Corresponding synthetic CT stacks were used to train the neural network in Biomedisa

volumes (Table 2). This observation is in accordance with a previous study (Rathnayaka et al. 2011), in which the use of several thresholds, that were determined locally for different parts of a bone, resulted in a higher accuracy of derived surfaces than global thresholding and Canny edge detection. Yet, the data presented herein show that the watershed segmentation, the weighted random walk approach, and the neural network produced slightly better segmentation results than automatic local thresholding for some of synthetic volumes. This mainly applies to volumes with low CNRs. Assuming that CT scanners are operated with reasonable settings, such low CNRs will rarely occur in reality. Thus, automatic local thresholding using the method *Otsu* with case-specific optimal parameters for the histogram window and radius can generally be recommended.

The choices of image gradients and seed regions had considerable effects on the segmentation quality of the watershed approach (Supporting Information Table S3). The gradients calculated by the *Gradient (3D)* function in Fiji and the *Sobel Filter* (applied in 3D) in Amira® performed comparably good and outperformed the other gradient calculation

methods with regard to the segmentation quality. In contrast, previous studies reported that Canny edge detection outperformed other edge detection or gradient calculation approaches for CT volumes (Rathnayaka et al. 2011; Jang et al. 2014) and non-medical images (Shin et al. 2001). These contradictory observations highlight the importance of carefully selecting an algorithm for image gradient calculation if watershed segmentation should be performed.

For watershed segmentation, the use of bone seed regions close to the correct segmentation (Phantom-1-Seeds) resulted in the best segmentation quality and the quality decreased with less accurate seed regions. Given that the correct segmentation usually is unknown for real CT volumes, determining an optimal seed region will be challenging in actual studies and automatic local thresholding might be the better choice as it might be more reliable in resulting in a high segmentation quality.

The neural network performed better than the other approaches for two of the synthetic CT volumes that were used to train the network (Table 2). For the other volumes, the segmentation accuracy achieved by the neural network was comparable to the results of the other methods. Given this pattern, it seems questionable if the neural network approach would result in the optimal segmentation quality for real CT volumes with unknown ground truth segmentations that could be used for training.

Surface generation and accuracy

Surfaces generated with the Amira® module *Generate Lego Surface* in combination with careful surface simplification (surface generation approach i) were the most accurate across voxels sizes (Fig. 2). This surface generation approach may, thus, be recommended. A maximal relative error of well below 0.4 voxels was obtained for almost all vertices in the surfaces; the observed error is possibly mostly due to the conversion of the continuous surface to a discrete volume. In absolute terms, the error increased with the voxel size and surfaces derived from the volume with the largest voxel size additionally contained a considerable number of artifacts (particularly in regions of thin bone; Fig. 2E). As these observations are in general accordance with previous studies (Fajardo et al. 2002; Gelaude et al. 2008; Kubo et al. 2008; Hassan et al. 2010; Christiansen, 2016), the use of surfaces derived from high resolution CT volumes seems recommendable if bone surfaces should be generated and measured.

In general, the accuracy of surfaces derived from the synthetic CT volumes by applying approach i decreased with the CNR and with the quality of the underlying segmentation (Fig. 3). There, however, were some exceptions: The surfaces derived from a synthetic CT volume

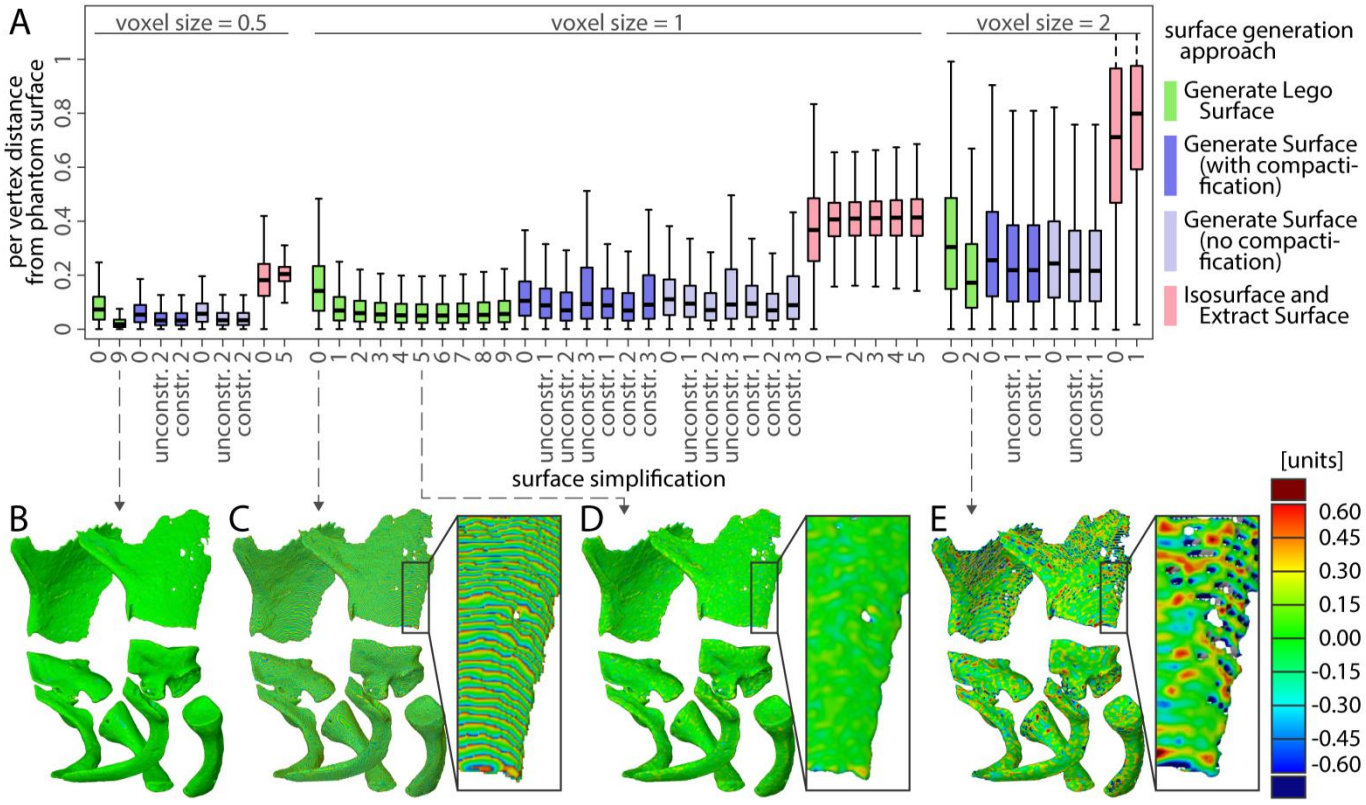


Fig. 2 Surface accuracies across voxel sizes. (A) Per vertex distance calculated from phantom surface to surfaces derived from error-free segmentations of volumes with different voxel sizes. Surface simplification: number of iterations of polygon count reduction and smoothing for 'Generate Lego Surface' (approach i) and 'Isosurface and Extract Surface' (ii), parameters used for *Smoothing type* and *Smoothing Extent* for other (iii); 0: no simplification. Range of different smoothing parameters given for voxel size of one, non-simplified and optimally simplified cases plotted for other voxel sizes. Distances/voxel sizes given in consistent unit. Dotted whisker: whisker cut. Potential outliers not plotted. (B)-(E) Heat map visualizations of per vertex distances calculated from derived surfaces to phantom surface. Dashed arrows indicate correspondence of box plots and heat map visualizations.

containing salt-and-pepper-like Poisson noise with a CNR of 4.2 were more accurate than the surfaces derived from a volume with a higher CNR of 4.8 and containing bubble-like real noise, although the latter yielded a higher segmentation quality (Table 2). This might indicate that an entirely random noise (e.g., Poisson noise) affects the surface accuracy less than a systematic or structured noise.

The surfaces derived from the watershed segmentation results were less accurate than the surfaces derived from the results of the other segmentation approaches in the three cases compared herein (Fig. 3A). Given that the watershed approach yielded a comparably high segmentation quality in those cases, watershed segmentation might not be best choice if accurate surfaces should be generated. Automatic local thresholding and, for some volumes also the segmentation methods implemented in Biomedisa are better choices to obtain a segmentation that allows for the generation of accurate surfaces.

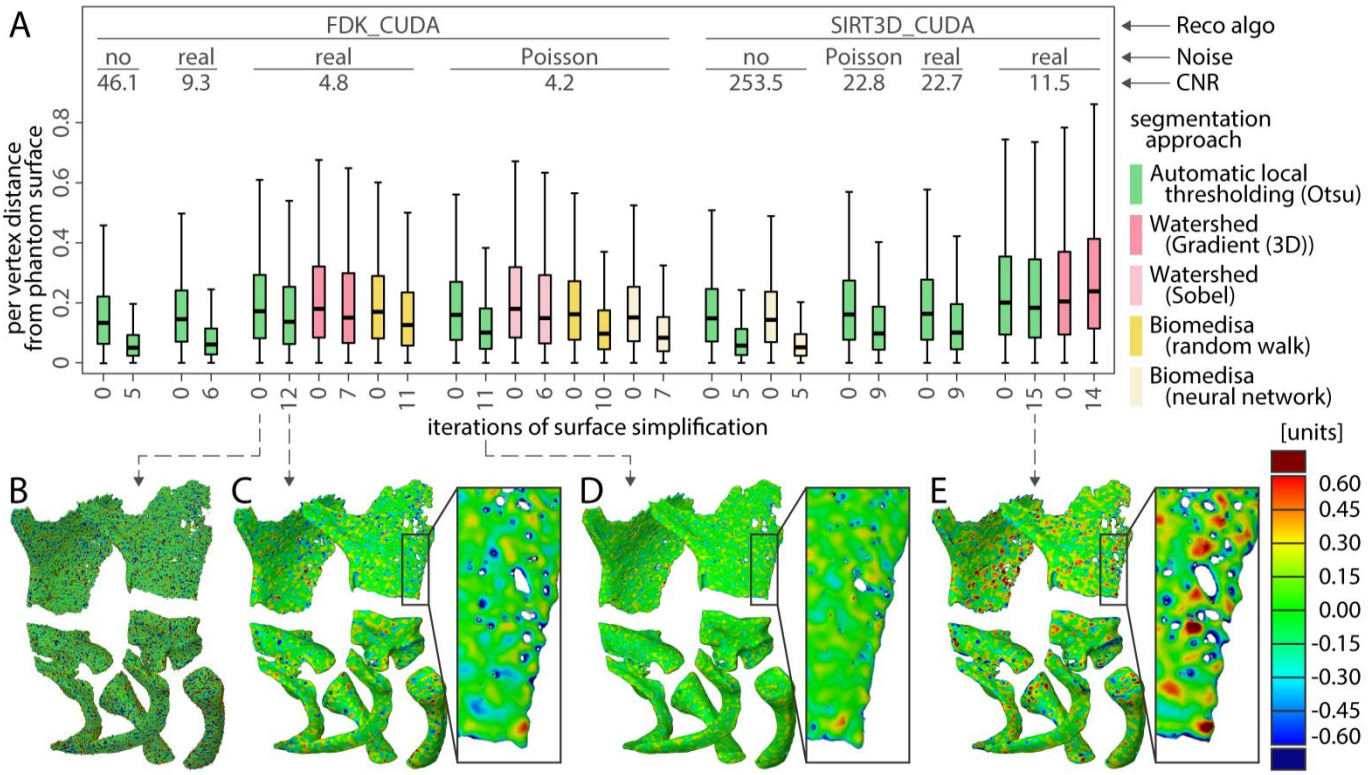


Fig. 3 Surface accuracy for different combinations of synthetic CT volumes and segmentation approaches. Automatic local thresholding performed on cross sections of the specimen. Surfaces generated by *Generate Lego Surface* module and simplified in Amira® (approach i). One unit equals one voxel. (A) Per vertex distance calculated from phantom surface to surfaces derived from best segmentation result obtained by respective segmentation approach. Iterations of surface simplification: number of iterations of polygon count reduction and smoothing; 0: no simplification; non-zero: optimally simplified case. Reco algo: reconstruction algorithm used during generation of synthetic CT volumes; Noise: type of noise added to the synthetic CT volumes; CNR: contrast to noise ratio calculated for bone and soft tissue. Potential outliers not plotted. (B)-(E) Heat map visualizations of per vertex distances calculated from derived surfaces to phantom surface. Dashed arrows indicate the correspondences of box plots and heat map visualizations.

The accuracy of the surfaces steadily increased with the degree of surface simplification up to a certain point and decreased from there on, but there was one exception (Fig. 3A): The first iteration of surface simplification decreased the accuracy of the surfaces derived from a synthetic CT volume with a low CNR (11.5) and a long gray gradient between structures; further simplification increased the accuracy again up to a certain degree of optimal simplification. In this particular case, the optimally simplified surface derived from the watershed segmentation result was less accurate than the corresponding unsimplified surface. This indicates that additional factors (possibly pattern of distribution of misclassified voxels, potential systematic deviations from the correct segmentation, ...) besides the segmentation quality might determine whether the effect of surface simplification is positive or negative. Nevertheless, surface simplification to an subjectively optimal degree seems recommendable

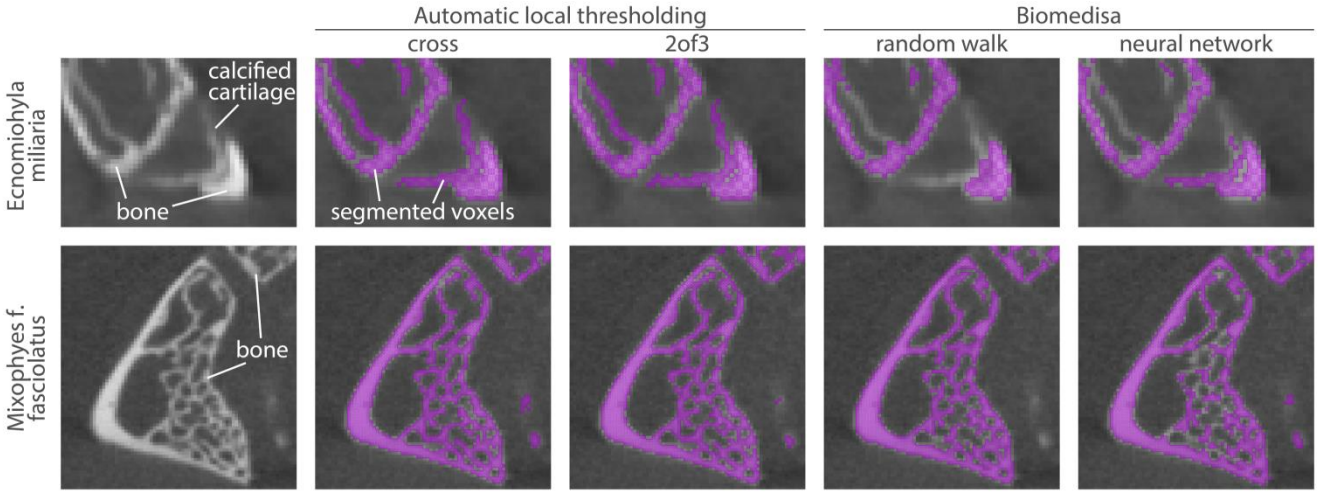


Fig. 4 Segmentation accuracies of pectoral girdle bones in real µCT volumes. Details of the CT volumes are showing bone adjacent to calcified cartilage (*Ecnomiohyla miliaria*) and trabecular bone (*Mixophyes fasciolatus fasciolatus*). Voxels segmented by the respective approaches are highlighted (violet).

as previous studies (Bade et al. 2006; DeVries et al. 2008; Veneziano et al. 2018) reported similar positive effects and surface simplification increases the accuracy with which landmark coordinates could be measured (Engelkes et al. 2019).

Comparison of segmentation approaches for real CT volumes

A visual comparison of the segmentation results and derived surfaces of real µCT volumes revealed that automatic local thresholding performed best in correctly classifying bone voxels of thin cortical and trabecular bone (1-3 layers of voxels; Figs 4, 5, Supporting Information Fig. S7). The second best segmentation results were obtained by the functions implemented in Biomedisa. Neither of the segmentation approaches compared herein was able to separate adjacent structures with similar gray values (e.g., bone and calcified cartilage); here, a manual separation of the structures was necessary. These observations generally support the findings derived from the synthetic CT volumes.

Cortical and trabecular bone are known to differ with regard to their mineral density (Gong et al. 1964) and trabeculae are represented by comparably low gray values in CT volumes (e.g., Suttapreyasri et al. 2018; Azhari et al. 2019). This likely imposes specific challenges for the segmentation of these structures and, in fact, the reported data show differences in the reliability with which trabeculae have been classified as bone (Fig. 4, Supporting Information Fig. S7). The visually best segmentation results were obtained with automatic local thresholding (cross and 2of3): trabeculae were, with some exceptions, satisfactorily segmented, although the respectively selected histogram window center possibly was not optimal for these structures. Given that using different thresholds for segmenting

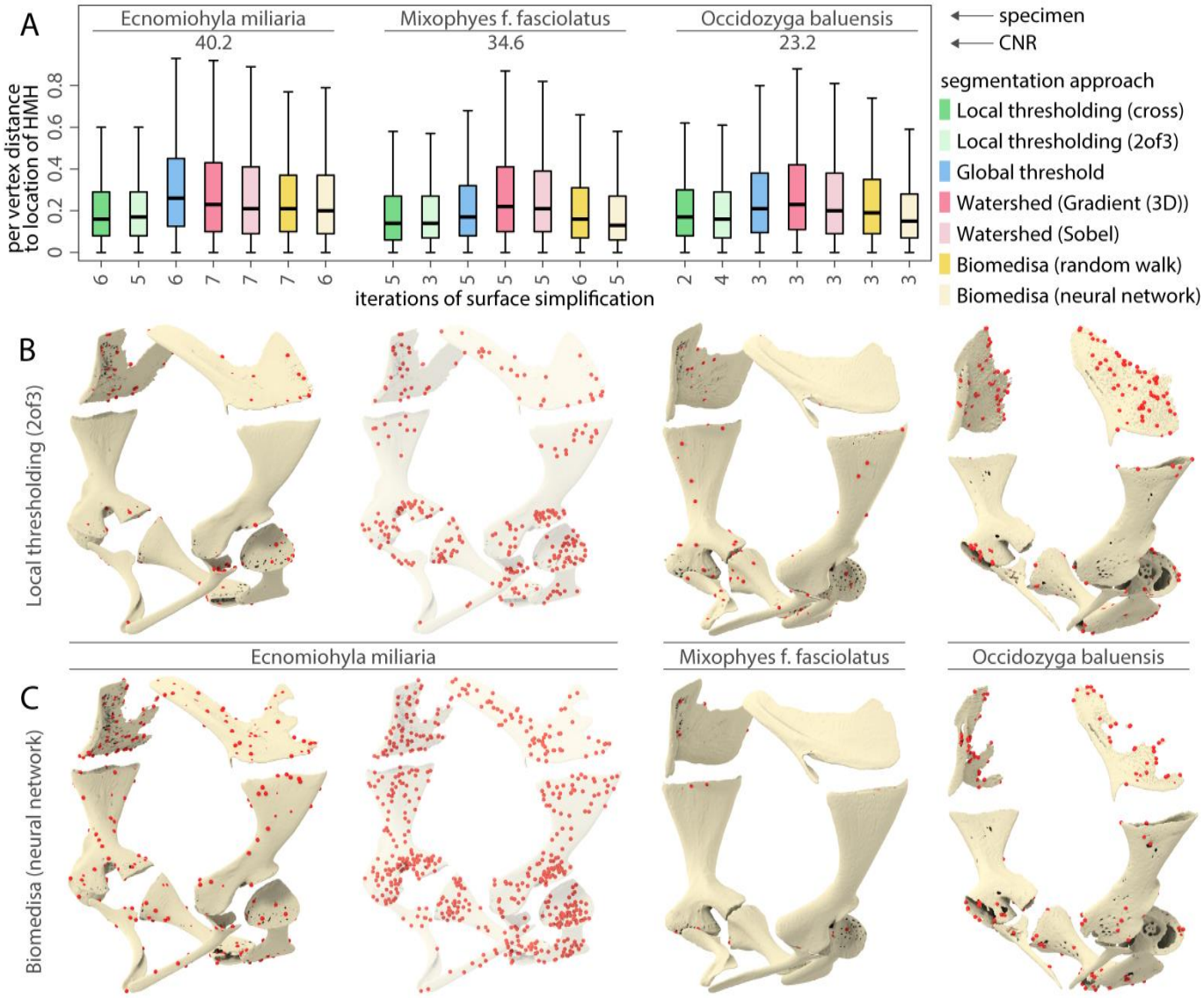


Fig. 5 Estimated surface accuracies of pectoral girdle bones in real µCT volumes. (A) Estimated accuracy of simplified surfaces derived from different segmentations of the µCT volumes; measured as the distance of selected vertices to the position of the half maximum height (HMH) along the respective surface normals. Note that this estimation is prone to different kinds of errors. CNR: contrast to noise ratio calculated for bone and soft tissue; local thresholding (cross/2of3): surfaces derived from the segmentation result of automatic local thresholding using the cross/2of3 approach; iterations of surface simplification: number of iterations of polygon count reduction and smoothing. One unit equals one voxel. Potential outliers not plotted. (B) Simplified surfaces of the segmentation results generated by automatic local thresholding with the 2of3 approach. Selected vertices that deviate more than 0.75 voxels from the position of the HMH highlighted by red spheres. Right surfaces of *Ecnomiohyla miliaria* rendered transparent. (C) Same as (B) but underlying segmentations generated by the trained neuronal network in Biomedisa.

cortical and trabecular bone has been shown to be superior to using a single threshold (Fajardo et al. 2002), the segmentation of trabecular bone obtained herein might have been improved by optimizing the window center for this specific structure. The weighted random walk implemented in Biomedisa visually performed second-best in segmenting trabecular

bone, followed by the performance of the neural network and global thresholding. It should be noted that the neural network used herein was trained for the synthetic volumes that contained considerably less trabeculae than the real volumes; a neural network specifically trained to segment cortical and trabecular bone might have perform better.

Considering the observed pattern of segmentation and surface qualities, automatic local thresholding with appropriately chosen parameters seems to be the most reliable approach for achieving high segmentation accuracies and may thus be recommended. Yet, the segmentation approaches implemented in Biomedisa are worth to be considered, in particular, because they require less volume pre-processing and are independent of the histogram window settings. In addition, neural networks have been reported to result in accurate bone segmentations for real CT volumes (e.g., Minnema et al. 2018; Vania et al. 2019).

Accuracy of surfaces derived from real CT volumes

The surfaces derived from the different segmentation results of the real CT volumes showed a considerable number of artifacts in areas of thin bone (1-2 layers of voxels), while surface areas of thicker bone could be considered satisfactory good (Fig. 5, Supporting Information Fig. S7). This highlights the importance of using a sufficient scan resolution, such that the bone parts that are of interest for measurement acquisition can be correctly segmented and converted to accurate surfaces.

Vertices with the highest deviation from the position of the local HMH mainly occurred on surface areas with numerous artifacts and on the inside of bones at trabeculae (Fig. 5B,C). The latter observation might be of particular importance for the usage of CT-derived surfaces in finite element (FE) analyses, as trabeculae are known to be a part of the load-transferring mechanism in bones (Barak et al. 2008) and their omission in FE models results in considerably higher von Mises stresses (Mielke & Nyakatura, 2019).

With regard to automatic local thresholding, a visual inspection revealed that surface areas parallel to the plane of thresholding tended to be noisy if the underlying segmentation was performed in one plane (cross). This effect did not occur if surfaces were derived from the segmentations resulting from the combination of different thresholding directions (2of3; Supporting Information Fig. S8). Given that the segmentation quality decreased to only a minor extent if the 2of3 approach was applied (Supporting Information Fig. S7, Table S2), the 2of3 approach can generally be recommended.

Quantitatively, the maximal per vertex deviation from the position of the local HMH was less than one voxel (with some outliers) for all derived surfaces. The accuracies of the

surfaces derived from the segmentations obtained by automatic local thresholding and by the neural network were similar (with the exception of *Ecnomiohyla miliaria*) and higher than the accuracies achieved by the other segmentation approaches. It should be kept in mind, that the used measure of the surface accuracy is prone to errors and possibly not suitable to quantify small differences. Yet, the observed accuracies of the surfaces derived from the segmentations by automatic local thresholding, and in two of the three cases also by the neural network, give confidence, that the underlying segmentations are reasonably good for those regions, in which bone voxels were classified as such.

Potential optimizations prior to segmentation

The reported data show that the quality (i.e., the CNR) and the voxel size of CT volumes have considerable effects on the accuracy of the segmentation and derived surface. Thus, it might be worth to spend some extra time on optimizing the steps of CT scanning and image processing prior to the segmentation process.

The determination of the correct boundary position within the gray scale gradient between adjacent structures is the main challenge in achieving an accurate segmentation of CT volumes. The gray scale gradient arises, among others, from the x-ray projections being blurred by the CT-scanner-specific point-spread function (PSF; Witt et al. 2003) and from the volume being additionally unsharpened by the reconstruction PSF that characterizes the blurring-effect of the backprojection (Orrison & Sanders, 1995; Buzug, 2008). The size (i.e., the full width a half maximum, FWHM) of the resulting PSF limits the spatial resolution of CT volumes in that the thickness of structures can only be accurately measured if the structures are thicker than the FWHM of the PSF (Dougherty & Newman, 1999; Prevrhal et al. 1999).

Different techniques to reduce the blurring in CT volumes have been proposed. The blur can be reduced by the deconvolution of the x-ray projections with the scanner-specific PSF; Wiener filtering needs to be performed during this process to limit the amplification of noise (Dougherty & Kawaf, 2001; Witt et al. 2003). In addition, the PSF was found to vary locally, which should be accounted for by using region-specific PSFs for the deconvolution (Witt et al. 2003). The focal-spot size of a x-ray tube has been observed to depended on the tube current and voltage (Chaney & Hendee, 1974) which renders the possibility likely that the PSF also changes with the power settings of the tube. Volumes reconstructed from deconvolved projections display sharper object boundaries and less variation in the gray values between objects of the same material but with different thicknesses (Witt et al. 2003).

A simple backprojection causes some blurring that is characterized by the PSF of the reconstruction process (Orrison & Sanders, 1995; Buzug, 2008; Hsieh, 2015). This PSF can be used to derive an image filter that counters the blurring effect in a filtered backprojection, but usually the results of this approach are sub-optimal (Orrison & Sanders, 1995). Another approach to reduce the blurring in a reconstructed CT volume was suggested by Pakdel et al. (2016): An estimation of a three-dimensional PSF can be derived from the reconstructed volume and may be used for iterative volume deconvolution; this approach leads to significantly more accurate quantitative bone measures (Pakdel et al. 2016). The three-dimensional PSF was found to be spatially variant for cone-beam CT (Chen & Ning, 2004), thus, the optimal volume deconvolution would likely be obtained if the spatial dependence of the PSF was accounted for.

Besides the volume blurring that can be characterized by a PSF, partial volume averaging non-linearly contributes to the gray scale gradient between adjacent structures (Goodenough et al. 1986; Fajardo et al. 2002; Barrett & Keat, 2004; Abel et al. 2013). Several approaches to reduce the effects of partial volume averaging have been published (e.g., Goodenough et al. 1986; Arabi & Kamali Asl, 2010; Heckel et al. 2014) and were recently reviewed in the context of positron emission tomography (Gargouri et al. 2018; Jomaa et al. 2018).

Although not addressed in the present study, CT volumes often contain artifacts like streaking, shading, or rings that presumably affect the segmentation and surface quality. These artifacts are caused by physical processes occurring during the CT data acquisition (e.g., beam hardening), by motions of the scanned object, or by sub-optimally functioning scanners (Barrett & Keat, 2004; Boas & Fleischmann, 2012). Securely placing the specimen within the field of view and deliberately choosing the scanner settings (i.e., appropriate number of projections, sufficient tube current and voltage, usage of a filter, correct calibration), as well as applying specific correction algorithms during the volume reconstruction might reduce such artifacts (Barrett & Keat, 2004; Boas & Fleischmann, 2012). Reducing such artifacts presumably will increase the quality of segmentations and derived surfaces, but this needs to be addressed in future studies.

The noise in the x-ray projection data can be reduced by longer exposure times and by averaging repeated frames acquired from the same angle (Keklikoglou et al. 2019), or by filtering the sinograms (Wang et al. 2005; Manduca et al. 2009). Noise in the CT volume can be reduced by increasing the number of projections (Keklikoglou *et al* 2019) or by using appropriately chosen iterative volume reconstruction algorithms (Renker et al. 2011; Boas &

Fleischmann, 2012; McLaughlin et al. 2014; Zhang & Xia, 2019). In addition, several approaches for reducing the noise in reconstructed volumes have been described (reviewed in Diwakar & Kumar, 2018) and image denoising has been reported to increase the volume quality (e.g., Kalra et al. 2003; Li et al. 2014; Vijaya & Suhasini, 2014). However, it is possible that image filtering enlarges (Diwakar & Kumar, 2018) or shifts the gray value gradient between structures, which in turn could affect the segmentation quality. Potential positive or negative effects of image denoising on the segmentation quality have rarely been assessed; positive effects have been reported for certain combinations of noise reduction methods and segmentation approaches (e.g., Firouzian et al. 2011; Nikonorov et al. 2016). Additionally, smoothing the segmentation label has been shown to increase the surface quality (DeVries et al. 2008). The effects of CT volume denoising or segmentation label smoothing on the segmentation quality of automatic local thresholding have, to the best of my knowledge, not been assessed yet.

Limitations

Most results reported in this study have been derived from synthetic CT volumes; the transferability of the recommendations to actual CT volumes was tested only superficially. The results need to be validated for more CT volumes and additional bone geometries.

The segmentation quality of automatic local thresholding considerably depended on the histogram window set prior to thresholding. This might have been an artifact caused by the conversion of the 16 bit volumes to 8 bit, which was necessary because the *Auto Local Threshold* plugin only accepts 8 bit images. The setting of a small histogram window prior to the conversion possibly resulted in retaining as much of the information relevant for separate bone from soft tissue voxels as possible. In this context, choosing a non-optimal value as center for the histogram window might have resulted in a shifted or somehow biased gray value gradient between bone and soft-tissue. This, however, is speculative and needs to be tested in future studies.

Neither segmentation approach tested in this study resulted in a satisfactory separation of adjacent structures with similar x-ray attenuation coefficients; human interaction is still needed to separate such structures. The segmentation strategies compared herein cover only a small subset of the approaches proposed in the literature (reviewed in Pham et al. 2000; van Eijnatten et al. 2018). Likewise, the surface generation and simplification strategies included in this study represent only some of the numerous approaches previously suggested (e.g.,

Bade et al. 2007; Moench et al. 2011; Gao et al. 2013; Chen et al. 2019). Those, herein neglected, approaches might result in higher segmentation or surface accuracies.

One of the herein neglected segmentation approaches that might be of particular interest for future studies is the application of automatic local thresholding in all three dimensions simultaneously instead of the slice-wise approach performed herein. Using a 3D patch (sub-volume) instead of a two-dimensional sub-image for the determination of a local threshold might have allowed for a more accurate threshold as information of more voxels in close proximity to the central voxel could have been considered.

Conclusion

High segmentation and surface accuracies can be obtained with readily available software packages. The highest achievable segmentation quality depends on the image quality of the underlying CT volume; segmentation results of images with a short gray value gradient between bone and soft tissue, a homogeneous gray value distribution within bones, and with a high CNR are most accurate. In most cases, automatic local thresholding with appropriately chosen parameters results in the best segmentation quality; the use of the method *Otsu* is recommendable. The center of the histogram window can be approximated from the mean gray value of bone and soft tissue. The window width should be as small as possible but wide enough such that no bone voxels disappear by getting black. The radius used for automatic local thresholding may be estimated from the CNR. The application of the 2of3 approach has only a minor effect on the thresholding quality but helps to reduce noise in the segmentation result and in derived polygon surfaces. Besides automatic local thresholding, the segmentation approaches implemented in Biomedisa (weighted random walk, neural network) are worth to be considered.

Surfaces derived from the segmentation results are most accurate if the surface generation algorithm implemented in the *Generate Lego Surface* module (Amira®) is used in combination with careful surface simplification. Surfaces with a maximal error of well below one voxel (except for some outliers) are obtainable even for volumes with comparably high image noise. These results need to be generalized and validated for more CT volumes. The herein presented strategies for assessing the accuracy of segmentations and surfaces can be adopted for future studies.

Acknowledgments

I thank Jörg U. Hammel and the Helmholtz-Zentrum Geesthacht for performing the µCT scan that was used to create the phantom surfaces, as well as Mehria Sedik for her support in gaining first experiences with assessing segmentation quality and Lena Kath for help with determining the CNRs and mean gray values of the real scans used in this study. Further, I am grateful to Jennice Helfsgott, Alexander Haas, and Frank Friedrich for discussing results of a pre-study, and to Philipp Lösel for help with Biomedisa. The author declares no conflict of interest. This study was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 387723284 to Alexander Haas.

Data Availability Statement

The phantom surface, phantom stack, and synthetic CT volumes can be downloaded from http://doi.org/10.25592/uhhfdm.962.

References

Abel RL, Prime M, Jin A, Cobb JP, Bhattacharya R (2013) 3D imaging bone quality: bench to bedside. *Hard Tissue* **2**, 42.

Andjelković M, Tomović L, Ivanović A (2016) Variation in skull size and shape of two snake species (*Natrix natrix* and *Natrix tessellata*). *Zoomorphology* **135**, 243–253.

Arabi H, Kamali Asl AR (2010) Feasibility study of a new approach for reducing of partial volume averaging artifact in CT scanner. In: *Proceedings of the 17th Iranian Conference of Biomedical Engineering (lCBME2010)*, pp. 1–4: IEEE.

Azhari, Oscandar F, Fariska I (2019) Normal value of cortical and mandibular trabecular bone density using cone beam computed tomography (CBCT). *J Int Dent Med Res* **12**, 160–164.

Bade R, Haase J, Preim B (2006) Comparison of fundamental mesh smoothing algorithms for medical surface models. In: *Simulation and Visualization* (eds Schulze T, Schlechtweg S, Hinz V), pp. 289–304. Erlangen: SCS European Publishing House.

Bade R, Konrad O, Preim B (2007) Reducing artifacts in surface meshes extracted from binary volumes. *J WSCG (Plzen)* **15**, 67–74.

Baillard C, Barillot C (2000) Robust 3D segmentation of anatomical structures with level sets. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000. Lecture Notes in Computer Science* (eds Goos G, Hartmanis J, van Leeuwen J, Delp SL, DiGoia AM, Jaramaz B), pp. 236–245. Berlin, Heidelberg: Springer Berlin Heidelberg.

Barak MM, Weiner S, Shahar R (2008) Importance of the integrity of trabecular bone to the relationship between load and deformation of rat femora: an optical metrology study. *J Mater Chem* **18**, 3855–3864.

Barandiaran I, Macía I, Berckmann E, et al. (2009) An automatic segmentation and reconstruction of mandibular structures from CT-data. In: *Intelligent Data Engineering*

and Automated Learning - IDEAL 2009. Lecture Notes in Computer Science (eds Corchado E, Yin H), pp. 649–655. Berlin, Heidelberg: Springer Berlin Heidelberg.

Barrett JF, Keat N (2004) Artifacts in CT: recognition and avoidance. *RadioGraphics* **24**, 1679–1691.

Baxter BS, Sorenson JA (1981) Factors affecting the measurement of size and CT number in computed tomography. *Invest Radiol* **16**, 337–341.

Boas FE, Fleischmann D (2012) CT artifacts: causes and reduction techniques. *Imaging Med* **4**, 229–240.

Bouxsein ML, Boyd SK, Christiansen BA, Guldberg RE, Jepsen KJ, Müller R (2010) Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. *J Bone Miner Res* **25**, 1468–1486.

Buzug T (2008) *Computed tomography: From Photon Statistics to Modern Cone-Beam CT*. Berlin, Heidelberg: Springer-Verlag. Available from: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10284881.

Chaney EL, Hendee WR (1974) Effects of x-ray tube current and voltage on effective focalspot size. *Med. Phys.* **1**, 141–147.

Chen Z, Hou Z, Xia M, Xing Y (2019) Adaptive surface mesh reconstruction for computed tomography images. *J Med Imaging Health Inform* **9**, 1086–1094.

Chen Z, Ning R (2004) Three-dimensional point spread function measurement of cone-beam computed tomography system by iterative edge-blurring algorithm. *Phys Med Biol* **49**, 1865–1880.

Christiansen BA (2016) Effect of micro-computed tomography voxel size and segmentation method on trabecular bone microstructure measures in mice. *Bone Rep* **5**, 136–140.

Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G (2008) MeshLab: an open-source mesh processing tool. In: *Sixth Eurographics Italian Chapter conference: Salerno, Italy, July 2nd - 4th, 2008. Eurographics Italian Chapter proceedings* (ed Scarano V), pp. 129–136. Aire-la-Ville: Eurographics Assoc.

Coleman MN, Colbert MW (2007) Technical note: CT thresholding protocols for taking measurements on three-dimensional models. *Am J Phys Anthropol* **133**, 723–725.

Cornette R, Baylac M, Souter T, Herrel A (2013) Does shape co-variation between the skull and the mandible have functional consequences? A 3D approach for a 3D problem. *J Anat* **223**, 329–336.

DeVries NA, Gassman EE, Kallemeyn NA, Shivanna KH, Magnotta VA, Grosland NM (2008) Validation of phalanx bone three-dimensional surface segmentation from computed tomography images using laser scanning. *Skeletal Radiol* **37**, 35–42.

Diwakar M, Kumar M (2018) A review on CT image noise and its denoising. *Biomed Signal Process Control* **42**, 73–88.

Dougherty G, Kawaf Z (2001) The point spread function revisited: image restoration using 2-D deconvolution. *Radiography* **7**, 255–262.

Dougherty G, Newman D (1999) Measurement of thickness and density of thin structures by computed tomography: a simulation study. *Med. Phys.* **26**, 1341–1348.

Engelkes K (2020) Phantom surface, phantom stack, synthetic CT volumes; 2020. Available from: http://doi.org/10.25592/uhhfdm.962.

Engelkes K, Friedrich F, Hammel JU, Haas A (2018) A simple setup for episcopic microtomy and a digital image processing workflow to acquire high-quality volume data and 3D surface models of small vertebrates. *Zoomorphology* **137**, 213–228.

Engelkes K, Helfsgott J, Hammel JU, et al. (2019) Measurement error in µCT-based three-dimensional geometric morphometrics introduced by surface generation and landmark data acquisition. *J Anat* **235**, 357–378.

Fajardo RJ, Ryan TM, Kappelman J (2002) Assessing the accuracy of high-resolution X-ray computed tomography of primate trabecular bone by comparisons with histological sections. *Am J Phys Anthropol* **118**, 1–10.

Firouzian A, Manniesing R, Flach ZH, et al. (2011) Intracranial aneurysm segmentation in 3D CT angiography: method and quantitative validation with and without prior noise filtering. *Eur J Radiol* **79**, 299–304.

Gao Z, Yu Z, Holst M (2013) Feature-preserving surface mesh smoothing via suboptimal Delaunay triangulation. *Graph Models* **75**, 23–38.

Gargouri S, Mouelhi A, Sayadi M, Labidi S, Mahersi M, Zayed S (2018) Brief review of partial volume corrections in PET/CT for lung cancer assessment. In: *4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6: IEEE.

Gelaude F, Vander Sloten J, Lauwers B (2006) Semi-automated segmentation and visualisation of outer bone cortex from medical images. *Comput Methods Biomech Biomed Engin* **9**, 65–77.

Gelaude F, Vander Sloten J, Lauwers B (2008) Accuracy assessment of CT-based outer surface femur meshes. *Comput Aided Surg* **13**, 188–199.

Gong JK, Arnold JS, Cohn SH (1964) Composition of trabecular and cortical bone. *Anat Rec* **149**, 325–331.

Goodenough DJ, Weaver KE, Costaridou H, Eerdmans H, Huysmans P (1986) A new software correction approach to volume averaging artifacts in CT. *Comput Radiol* **10**, 87–98.

Guyomarc'h P, Santos F, Dutailly B, Desbarats P, Bou C, Coqueugniot H (2012) Three-dimensional computer-assisted craniometrics: a comparison of the uncertainty in measurement induced by surface reconstruction performed by two computer programs. *Forensic Sci Int* **219**, 221–227.

Hassan B, Couto Souza P, Jacobs R, Azambuja Berti S de, **van der Stelt P** (2010) Influence of scanning and reconstruction parameters on quality of three-dimensional surface models of the dental arches from cone beam computed tomography. *Clin Oral Investig* **14**, 303–310.

Healy S, McMahon J, Owens P, Dockery P, FitzGerald U (2018) Threshold-based segmentation of fluorescent and chromogenic images of microglia, astrocytes and oligodendrocytes in FIJI. *J Neurosci Meth* **295**, 87–103.

Heckel F, Meine H, Moltz JH, et al. (2014) Segmentation-based partial volume correction for volume estimation of solid lesions in CT. *IEEE Trans. Med. Imaging* **33**, 462–480.

Hsieh J (2015) *Computed tomography: principles, design, artifacts, and recent advances*. Bellingham: Society of Photo-Optical Instrumentation Engineers (SPIE).

Ito T (2019) Effects of different segmentation methods on geometric morphometric data collection from primate skulls. *Methods Ecol Evol* **10**, 1972–1984.

Jang S-W, Seo Y-J, Yoo Y-S, Kim YS (2014) Computed tomographic image analysis based on FEM performance comparison of segmentation on knee joint reconstruction. *ScientificWorldJournal* **2014**, 235858.

Jomaa H, Mabrouk R, Khlifa N (2018) Post-reconstruction-based partial volume correction methods: a comprehensive review. *Biomed Signal Process Control* **46**, 131–144.

Kalra MK, Maher MM, Sahani DV, et al. (2003) Low-dose CT of the abdomen: evaluation of image improvement with use of noise reduction filters - pilot study. *Radiology* **228**, 251–256.

Keklikoglou K, Faulwetter S, Chatzinikolaou E, et al. (2019) Micro-computed tomography for natural history specimens: a handbook of best practice protocols. *Eur J Taxon* **522**, 1–55.

Koehler PR, Anderson RE, Baxter B (1979) The effect of computed tomography viewer controls on anatomical measurements. *Radiology* **130**, 189–194.

Kubo D, Kono RT, Saso A, Mizushima S, Suwa GEN (2008) Accuracy and precision of CT-based endocranial capacity estimations: a comparison with the conventional millet seed method and application to the Minatogawa 1 skull. *Anthropol Sci* **116**, 77–85.

Lamecker H, Seebass M, Hege H-C, Deuflhard P (2004) A 3D statistical shape model of the pelvic bone for segmentation. In: *Medical Imaging 2004: Image Processing. SPIE Proceedings* (eds Fitzpatrick JM, Sonka M), pp. 1341–1351: SPIE.

Landini G, Randell DA, Fouad S, Galton A (2017) Automatic thresholding from the gradients of region boundaries. *J Microsc* **265**, 185–195.

Li Z, Yu L, Trzasko JD, et al. (2014) Adaptive nonlocal means filtering based on local noise level for CT denoising. *Med. Phys.* **41**, 11908.

Lindh C, Obrant K, Petersson A (2004) Maxillary bone mineral density and its relationship to the bone mineral density of the lumbar spine and hip. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **98**, 102–109.

Lösel P, Heuveline V (2016) Enhancing a diffusion algorithm for 4D image segmentation using local information. In: *Medical Imaging 2016: Image Processing. SPIE Proceedings* (eds Styner MA, Angelini ED), 97842L: SPIE.

Magnusson A (1987) Object size determination at computed tomography. *Upsala J Med Sci* **92**, 277–286.

Manduca A, Yu L, Trzasko JD, et al. (2009) Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med. Phys.* **36**, 4911–4919.

McLaughlin PD, Murphy KP, Hayes SA, et al. (2014) Non-contrast CT at comparable dose to an abdominal radiograph in patients with acute renal colic; impact of iterative reconstruction on image quality and diagnostic performance. *Insights into imaging* **5**, 217–230.

Mielke M, Nyakatura JA (2019) Bone microstructure in finite element modeling: the functional role of trabeculae in the femoral head of *Sciurus vulgaris*. *Zoomorphology* **138**, 535–547.

Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J (2018) CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput Biol Med* **103**, 130–139.

Mitchell HB (2010) *Image Fusion: Theories, Techniques and Applications*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Moench T, Gasteiger R, Janiga G, Theisel H, Preim B (2011) Context-aware mesh smoothing for biomedical applications. *Comput. Graph.* **35**, 755–767.

Nikonorov A, Kolsanov A, Petrov M, et al. (2016) Vessel segmentation for noisy CT data with quality measure based on single-point contrast-to-noise ratio. In: *E-Business and Telecommunications. Communications in Computer and Information Science* (eds Obaidat MS, Lorenz P), pp. 490–507. Cham: Springer International Publishing.

Orrison WW, Sanders JA (1995) Clinical brain imaging: computerized axial tomography and magnetic resonance imaging. In: *Functional Brain Imaging*, pp. 97–144: Elsevier.

Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* **9**, 62–66.

Pakdel A, Hardisty M, Fialkov J, Whyne C (2016) Restoration of thickness, density, and volume for highly blurred thin cortical bones in clinical CT images. *Ann Biomed Eng* **44**, 3359–3371.

Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. *Annu Rev Biomed Eng* **2**, 315–337.

Phansalkar N, More S, Sabale A, Joshi M (2011) Adaptive local thresholding for detection of nuclei in diversity stained cytology images. In: *International Conference on Communications and Signal Processing (ICCSP)*, pp. 218–220: IEEE.

Prevrhal S, Engelke K, Kalender WA (1999) Accuracy limits for the determination of cortical width and density: the influence of object size and CT imaging parameters. *Phys Med Biol* **44**, 751–764.

R Core Team (2019) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Rathnayaka K, Sahama T, Schuetz MA, Schmutz B (2011) Effects of CT image segmentation methods on the accuracy of long bone 3D reconstructions. *Med Eng Phy* **33**, 226–233.

Renker M, Nance JW, Schoepf UJ, et al. (2011) Evaluation of heavily calcified vessels with coronary CT angiography: comparison of iterative and filtered back projection image reconstruction. *Radiology* **260**, 390–399.

Roschger P, Paschalis EP, Fratzl P, Klaushofer K (2008) Bone mineralization density distribution in health and disease. *Bone* **42**, 456–466.

RStudio Team (2018) *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. Available from: http://www.rstudio.com/.

Scherf H, Tilgner R (2009) A new high-resolution computed tomography (CT) segmentation method for trabecular bone architectural analysis. *Am J Phys Anthropol* **140**, 39–51.

Schindelin J, Arganda-Carreras I, Frise E, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675.

Seibert CE, Barnes JE, Dreisbach JN, Swanson WB, Heck RJ (1981) Accurate CT measurement of the spinal cord using metrizamide: physical factors. *Am J Roentgenol* **136**, 777–780.

Shin MC, Goldgof DB, Bowyer KW, Nikiforou S (2001) Comparison of edge detection algorithms using a structure from motion task. *IEEE Trans Syst Man Cybern B Cybern* **31**, 589–601.

Spoor CF, Zonneveld FW, Macho GA (1993) Linear measurements of cortical bone and dental enamel by computed tomography: applications and problems. *Am J Phys Anthropol* **91**, 469–484.

Suttapreyasri S, Suapear P, Leepong N (2018) The accuracy of cone-beam computed tomography for evaluating bone density and cortical bone thickness at the implant site: micro-computed tomography and histologic analysis. *J Craniofac Surg* **29**, 2026–2031.

Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* **15**, 29.

Ullrich CG, Binet EF, Sanecki MG, Kieffer SA (1980) Quantitative assessment of the lumbar spinal canal by computed tomography. *Radiology* **134**, 137–143.

van Aarle W, Palenstijn WJ, Beenhouwer J de**, et al.** (2015) The ASTRA toolbox: a platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **157**, 35–47.

van Aarle W, Palenstijn WJ, Cant J, et al. (2016) Fast and flexible X-ray tomography using the ASTRA toolbox. *Opt Express* **24**, 25129–25147.

van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J (2018) CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phy* **51**, 6–16.

Vania M, Mureja D, Lee D (2019) Automatic spine segmentation from CT images using Convolutional Neural Network via redundant generation of class labels. *J Comput Des Eng* **6**, 224–232.

Veneziano A, Landi F, Profico A (2018) Surface smoothing, decimation, and their effects on 3D biological specimens. *Am J Phys Anthropol* **166**, 473–480.

Vijaya G, Suhasini A (2014) An adaptive preprocessing of lung CT images with various filters for better enhancement. *Academic Journal of Cancer Research* **7**, 179–184.

Vincent L, Soille P (1991) Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Trans Pattern Anal Machine Intell* **13**, 583–598.

Wang J, Lu H, Li T, Liang Z (2005) Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters. In: *Medical Imaging 2005: Image Processing. SPIE Proceedings* (eds Fitzpatrick JM, Reinhardt JM), p. 2058: SPIE.

Wickham H (2016) *Ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag. Available from: http://ggplot2.org.

Witt ST, Riedel CH, Goessl M, Chmelik MS, Ritman EL (2003) Point spread function deconvolution in 3D micro-CT angiography for multiscale vascular tree separation. In: *Medical Imaging 2003: Physics of Medical Imaging. SPIE Proceedings* (eds Yaffe MJ, Antonuk LE), p. 720: SPIE.

Zhang S, Xia Y (2019) CT image reconstruction algorithms: a comprehensive survey. *Concurrency Computat Pract Exper* **217**, 430.

Supporting Information

[The following Supporting Information can be found at the end of this thesis, pp. 244-254.]

Fig. S1 Cross section of µCT volume of *Bombina orientalis* (ZMH A12601) illustrating the characteristics of such data.

Fig. S2 Phantom surface and phantom stack.

Fig. S3 Workflow to generate synthetic CT volumes.

Fig. S4 Slices of selected synthetic CT volumes showing different lengths of the gray value gradient between structures, gray value distributions within bones, and noise characteristics.

Fig. S5 Workflow of performing segmentation trials to assess the performance (quality) of different approaches.

Fig. S6 Illustration of the accuracy assessment of a surface derived from real µCT volumes.

Fig. S7 Segmentation and surface accuracies of pectoral girdle bones in real µCT scans.

Fig. S8 Comparison of unsimplified surfaces of *Ecnomiohyla miliaria* derived from different segmentation results obtained by automatic local thresholding.

Table S1 MicroCT volumes downloaded from MorphoSource and used in this study.

Table S2 Highest achieved segmentation quality of automatic local and global thresholding and respective parameters used to obtain best thresholding results.

Table S3 Image-specific segmentation quality achieved with different approaches of watershed segmentation; quality measured as percentage of correctly classified voxels at the border of the bones ±2 layers of voxels.

- Chapter four -

Ecomorphology of the pectoral girdle in anurans (Amphibia, Anura): shape diversity and biomechanical considerations

Karolin Engelkes[1], Lena Kath[1], Thomas Kleinteich[2], Jörg U. Hammel[3,4], André Beerlink[5], Alexander Haas[1]

1) Center of Natural History (CeNak), Universität Hamburg, Hamburg, Germany
2) TPW Prüfzentrum GmbH, Neuss, Germany
3) Institute of Materials Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany
4) Institut für Zoologie und Evolutionsforschung mit Phyletischem Museum, Ernst-Hackel-Haus und Biologiedidaktik, Friedrich-Schiller-Universität Jena, Jena, Germany
5) YXLON International GmbH, Hamburg, Germany

Published in Ecology and Evolution, "Early View", doi 10.1002/ece3.6784.

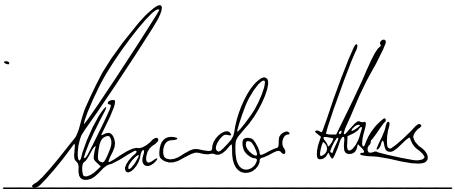
Contributions of KE:

Study design: 100%

Data collection: CT scanning ~ 70%, segmentation & surface generation ~ 95%, landmark data acquisition 100%, musculoskeletal modelling 100%, FEA 100%

Data analysis: 100%

Manuscript drafting: 100%

Contributions confirmed by supervisor: 

Prof. Dr. Alexander Haas

Abstract

Frogs and toads (Lissamphibia: Anura) show a diversity of locomotor modes that allow them to inhabit a wide range of habitats. The different locomotor modes are likely to be linked to anatomical specializations of the skeleton within the typical frog Bauplan. While such anatomical adaptations of the hind limbs and the pelvic girdle are comparably well understood, the pectoral girdle received much less attention in the past. We tested for locomotor-mode-related shape differences in the pectoral girdle bones of 64 anuran species by means of micro-computed-tomography-based geometric morphometrics. The pectoral girdles of selected species were analyzed with regard to the effects of shape differences on muscle moment arms across the shoulder joint and stress dissipation within the coracoid. Phylogenetic relationships, size, and locomotor behavior have an effect on the shape of the pectoral girdle in anurans, but there are differences in the relative impact of these factors between the bones of this skeletal unit. Remarkable shape diversity has been observed within locomotor groups indicating many-to-one mapping of form onto function. Significant shape differences have mainly been related to the overall pectoral girdle geometry and the shape of the coracoid. Most prominent shape differences have been found between burrowing and nonburrowing species with headfirst and backward burrowing species significantly differing from one another and from the other locomotor groups. The pectoral girdle shapes of burrowing species have generally larger moment arms for (simulated) humerus retractor muscles across the shoulder joint, which might be an adaptation to the burrowing behavior. The mechanisms of how the moment arms were enlarged differed between species and were associated with differences in the reaction of the coracoid to simulated loading by physiologically relevant forces.

Keywords: Locomotion; many-to-one mapping; muscle moment arm; stress dissipation; trade-off

Introduction

Frogs and toads (Lissamphibia: Anura) inhabit a wide range of habitats and, among other things, utilize different locomotor behaviors within those habitats (Wells, 2007). Almost all anurans are capable of some kind of hopping or jumping (Wells, 2007) and swimming (Abourachid & Green, 1999). Some species have been reported to extensively use quadrupedal walking (e.g., Ahn, Furrow & Biewener, 2004); other, in particular, fossorial species show burrowing behavior by moving the substrate using either their hind legs, forelegs, or heads (e.g., Emerson, 1976; Nomura, Rossa-Feres, & Langeani, 2009). Arboreal species are able to climb in vegetation (Herrel et al., 2013); some of them have evolved parachuting or gliding abilities (Oliver, 1951).

From an ecomorphological point of view, different behaviors and the associated performances provide the link between the morphology of a specimen and its ecology (e.g., Arnold, 1983; Ricklefs & Miles, 1994; Wainwright, 1994). The anatomy determines the

functional properties, which in turn determine the performance capacities of a specimen (Wainwright, 2007). In this context, natural selection should favor anatomical peculiarities that allow high locomotor performances, as moving in space is crucial for individuals, for example, to use the resources of their habitat, to encounter mates, or to escape from predators (e.g., Liedvogel, Chapman, Muheim, & Åkesson, 2013; Nathan et al., 2008).

Previous studies (e.g., Citadini, Brandt, Williams, & Gomes, 2018; Emerson, 1988; Vera, Ferretti, Abdala & Cointry, 2020; Zug, 1978) have reported associations of anatomical traits with locomotor behavior or performance, or ecology within the Anura. Most of these studies have focused on the pelvic girdle, the relative length of fore- or hind limbs, or the anatomy of the hind limbs. For example, the takeoff speed for jumping was found to be associated with hind limb length, hind limb muscle mass, and muscle contraction rates (Choi & Park, 1996; Choi, Shim, & Ricklefs, 2003) and specific locomotor modes tended to be associated with differences in the shape of the sacrum (Jorgensen & Reilly, 2013).The mechanical properties of the femur and tibiofibula differed between certain locomotor modes (Vera et al., 2020; Wilson, Espinoza, Shah, & Blob, 2009;). Species inhabiting the same microhabitat were similar with regard to their hind limb morphology, external body proportions, and performance in selected ecologically relevant tasks (Moen, Irschick, & Wiens, 2013).

High jumping performance, for instance, was generally associated with relatively short forelimbs (Zug, 1972), comparably long hind limbs (e.g., Astley, 2016; Emerson, 1978) with tibiofibulae being longer than the femora (Gómez & Lires, 2019), larger hind limb muscles (e.g., Astley, 2016), and specific physiological muscle properties (e.g., Astley, 2016; Chadwell, Hartwell & Peters, 2002). The difference in the length of the hind limbs compared to the forelimbs was less pronounced in primary walking species (Reynaga, Astley & Azizi, 2018). Jumping and walking, hopping species have been reported to differ in the anatomy of the ilio-sacral joint and the associated configuration of the ilio-lumbaris muscle, although there were some exceptions in the correlation of joint anatomy with locomotor mode (Emerson, 1979).

In addition, previously recognized morphological adaptations to swimming involved specific relative limb proportions (Gómez & Lires, 2019) and extensive foot webbing (Laurent, 1964). Additionally, the relative muscle mass of the hind limbs in frequently swimming species was higher if compared to other species (Moen, 2019). The ilio-sacral joint in the aquatic species *Xenupus laevis* allowed for sliding and was thought to thereby increase

the length of the power stroke and to contribute to fast submerging after breathing (Videler & Jorna, 1985).

Climbing behavior was usually associated with a bicondylar sacro-urostylic articulation (Reilly & Jorgensen, 2011), large finger and toe tips (Moen et al., 2013), adhesive toe pads (Emerson & Diehl, 1980; Noble & Jaeckle, 1928), and modifications of the finger extensor muscles (Burton, 1998). In addition, hands and feet could be webbed (Laurent, 1964), the distal forelimbs of certain species might be adapted to grasping (Manzano, Abdala & Herrel, 2008), and the presence of an intercalary cartilage or bone between the two terminal phalanges in some arboreal anuran species was thought to increase the efficiency of the adhesive toe pads (Noble & Jaeckle, 1928).

Finally, the body of burrowing species was generally observed to be globular (Dutta & Pradhan ,1985; Laurent, 1964) with relatively shorter and stronger limbs (Laurent, 1964; Moen, 2019) and a short tibiofibula relative to the femur (Enriquez-Urzelai, Montori, Llorente & Kaliontzopoulou, 2015; Gómez & Lires, 2019). Most backward burrowing species had enlarged metatarsal tubercles (Kley & Kearney, 2006; Moen et al., 2013). Short hind limbs and the presence of metatarsal tubercles have been suggested to increase the performance of backward burrowing (Emerson, 1976). Further examples of the adaptation to backward burrowing include the increase in the size and robustness of the prehallux (Kley & Kearney, 2006) and species-specific modification of the feet muscles (Blotto, Pereyra, Faivovich, Dias, & Grant, 2017; Burton, 2001; Sanders & Davies, 1983). Headfirst burrowing has been reported to be species-specifically associated with a modified skull (Davies, 1984; Menzies & Tyler, 1977), massive mandibles (Menzies & Tyler, 1977), relatively short and robust forelimbs (Brown, Jackson & Brown, 1972), or modifications of the manus (Kley & Kearney, 2006).

The forelimbs of anurans have been reported to accomplish species- and case-specific tasks during locomotion (e.g., hopping/jumping: Nauwelaerts & Aerts, 2006; swimming: Abourachid & Green, 1999; Gillis & Biewener, 2000; walking: Reynaga et al., 2018; burrowing: Sanders & Davies, 1983; climbing: Manzano et al., 2008). The forelimbs, for example, decelerate the body during coordinated landing (Cox & Gillis, 2015), move the soil during headfirst burrowing (Emerson, 1976), or stabilize the body during gliding (Emerson & Koehl, 1990). In addition, some muscles originating from the pectoral girdle and inserting onto the forelimb have been shown to be active during different phases of a jump (Akella & Gillis, 2011). Yet, the pectoral girdle, that is, the central element linking the forelimbs to the

TABLE 1 Definition of locomotor groups.

Locomotor group	Definition
swimming	Purely aquatic locomotion.
walking, hopping	Quadrupedal walking or hopping (sensu Emerson 1979: jumps with a maximum length of less than 8-9 times snout-vent-length) on land. Optional swimming behavior, no climbing or burrowing.
jumping	Same as "walking, hopping" but with maximum jumps longer than 8-9 times snout-vent-length (Emerson 1979).
backward burrowing	Swimming, walking, hopping, or jumping but with additional digging using the hind limbs. No use of arms/head for digging.
headfirst burrowing	Same as "backward burrowing" but additional use of forelimbs or head to move soil.
climbing	Swimming, walking, hopping, or jumping but with additional climbing and jumping locomotion in vegetation. Optional parachuting or gliding locomotion (sensu Oliver 1951: while falling descending along path that deviates less [parachuting] or more [gliding] than 45° from the vertical).

axial skeleton, has received little attention regarding the association of anatomical traits with and the functional adaptation to different locomotor behaviors.

Different pectoral girdle types (arciferal, firmisternal) were suggested to accomplish similar tasks (i.e., dissipating landing forces), but in different ways (Emerson, 1983, 1984). One previous study reported that higher jumping abilities were associated with shorter scapulae (Zug, 1972), whereas another observed jumping species to have long scapulae with broad proximal and distal ends, and long claviculae and coracoids (Soliz, Tulli, & Abdala, 2017). Headfirst burrowing was associated with a forward shifted scapula causing the suprascapula to overlap the posterior margin of the skull, and robust and posteromedially directed coracoids in some species (Davies, 1984; Emerson, 1976). Besides these partly contradictory reports, little is known about the anuran pectoral girdle in relation to different locomotor behaviors and on the biomechanical functions of this skeletal complex during locomotion.

Here, we aim to resolve the relationships between locomotor mode, shape variation, and biomechanical function of the pectoral girdle of anurans. To do so, selected anuran species were assigned to one of six groups of locomotor behavior (subsequently called locomotor groups) and the shape of their pectoral girdle bones was assessed by means of geometric morphometrics. The phylogenetic signal was determined, and shape differences among locomotor groups were statistically assessed. The pectoral girdles of selected species were analyzed with regard to the effects of shape differences on muscle moment arms across the shoulder joint and simulated stress dissipation within the coracoid. Results were discussed in the context of adaptation to locomotor behaviors.

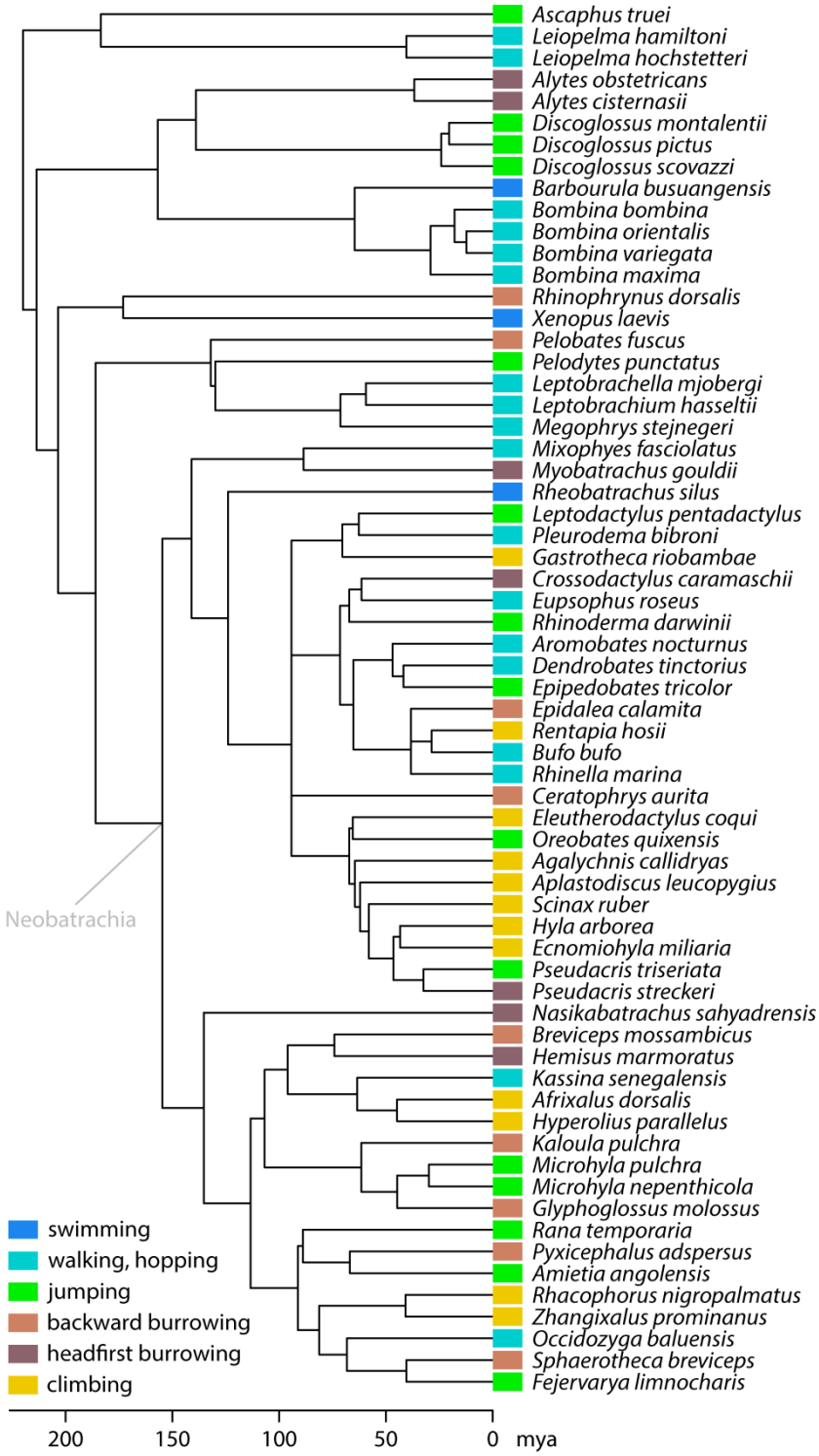


FIGURE 1. Phylogenetic relation and locomotor behavior of species examined in this study

Material and methods

Specimens and µCT scanning

Locomotor groups were defined (Table 1) and assigned based on literature accounts (Appendix S1: Tables A1, A2). Sixty-four species (Figure 1) covering 31 of the 52 currently recognized (Frost 2020) anuran (Amphibia: Anura) families were selected based on their

phylogenetic position and locomotor behavior. A time-calibrated phylogeny was extracted from TimeTree.org (accessed 2nd March 2020; Kumar, Stecher, Suleski, & Hedges, 2017); six species were replaced by close relatives (assessed from Pyron & Wiens, 2011) for extraction as they were not listed on TimeTree.org. Species names were updated following Frost (2020). The aim was to achieve heterogeneous subclades with regard to locomotor behavior and a wide dispersion of locomotor groups across the phylogeny in order to avoid potential negative effects on the statistical analyses (Adams & Collyer, 2018).

Selected micro-computed tomography (µCT) scans of a previous study (Engelkes et al., 2019) were used in combination with additional µCT volumes. Scans were performed with a Skyscan 1172 (Bruker microCT), Phoenix Nanotom S or M (GE Sensing & Inspection Technologies GmbH), Phoenix v|tome|x L 450 (GE Sensing & Inspection Technologies GmbH), or a YXLON FF20 CT or FF35 CT (YXLON International GmbH; Appendix S1: Table A1). Additional µCT volumes were downloaded from MorphoSource (https://www.morphosource.org/; Appendix S1: Table A2).

Segmentation and surface generation

A previous study (Engelkes et al., 2019) found that the techniques applied to generate the polygon surfaces have a significant effect on the landmark data acquired from them. The workflow herein followed the recommendations in Engelkes (preprint, in review) in order to obtain surfaces that are as accurate as possible. The pectoral girdle bones (including calcified sternal or episternal elements, if applicable) were roughly segmented in Amira (version 6.0.1; Konrad-Zuse-Zentrum Berlin, FEI Visualization Sciences Group), and the mean gray value *m* of pectoral girdle bones and surrounding soft tissues, and the standard deviation of the soft tissue gray values were determined for each original CT volume separately. The mean gray value *m* was used to set limits to the gray value histogram of the respective CT volume in Fiji (based on ImageJ 1.51n; Schindelin et al.; 2012; Schneider, Rasband, & Eliceiri, 2012). The limits were chosen such that they laid symmetrically around the value calculated by 1.019 * *m* - 462.812 (see Engelkes, in review for the derivation of this formula) and such that the contrast of bone and surrounding voxels was maximized without bone voxels getting black.

Each adjusted CT volume was resliced from top to bottom and from left to right, and all stacks were thresholded by automatic local thresholding (Fiji plugin *Auto Local Threshold*, Landini, Rueden, Schindelin, Hiner, & Helfrich, https://imagej.net/Auto_Local_Threshold). The three thresholding results of each CT volume were combined in Amira by setting those

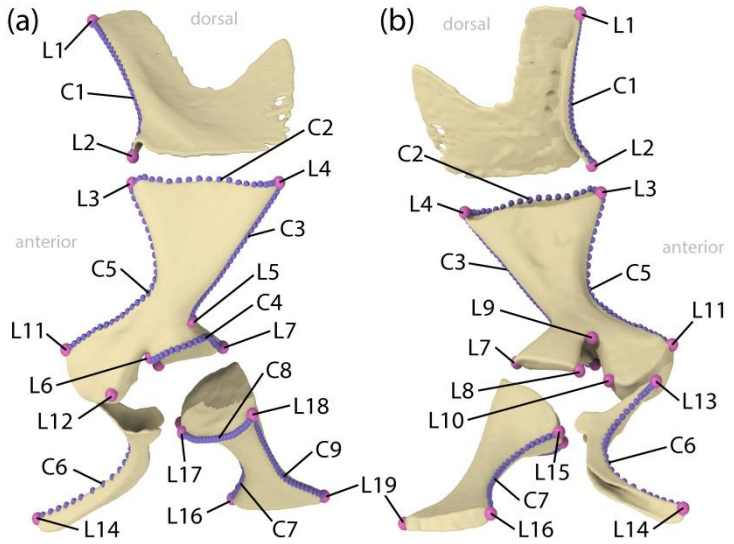


FIGURE 2. Landmarks (pink; L1-19) and semilandmarks (violet; C1-9) on the bones of the left-side pectoral girdle of *Ecnomiohyla miliaria*. (a) Lateral view. (b) Medial view

voxels as bone that were classified as bone in any two of the three thresholded stacks. The resulting stack was combined with the rough segmentation of the pectoral girdle bones to separate the bones form other structures. Foramina were filled and arteifacts (i.e., segmented noise, unsegmented bone voxels) were corrected in regions in which semilandmarks should be placed by manually adjusting the segmentation accordingly.

Polygon surfaces were generated using the *Generate Lego Surface* module in Amira in combination with surface simplification (reduction of polygon count and smoothing; *Simplification Editor* and *Smooth Surface* module) to a subjective optimal degree. Surface generation and simplification were accelerated by a modified version of the *MultiExport* macro (Engelkes, Friedrich, Hammel, & Haas, 2018). The bones of the right pectoral girdle halves were mirrored (MeshLab version 1.3.3; Cignoni et al., 2008) to the left to avoid any potential bias due to orientation during landmark acquisition. Surfaces with major deformations or artifacts were excluded from subsequent steps.

Landmarks and superimposition

Landmarks were, with slight modifications, adopted from Engelkes et al. (2019) and complemented by curves of sliding semilandmarks (Gunz & Mitteroecker, 2013; Figure 2; Appendix S1: Table A3). For each pectoral girdle half, 19 fixed landmarks (including start and end points of curves) and nine curves with 21 to 29 semilandmarks were acquired in Stratovan Checkpoint (version 2020.02.05.1043; Stratovan Corporation). No landmarks were acquired from the sternum or episternum, as those structures were present in only some species. Three microhylid species (*Kaloula pulchra*, *Microhyla nepenthicola*, and *M. pulchra*)

lacked a clavicula and, consequently, the (semi)landmarks on the clavicula were missing in the respective landmark configurations.

All subsequent steps were performed in R (version 3.5.3; R Core Team, 2019) using RStudio (version 1.1.463; RStudio Team, 2018) and functions of the packages abind (version 1.4-5; Plate & Heiberger, 2016), ape (version 5.3; Paradis & Schliep, 2018), geomorph (version 3.2.1; Adams, Collyer, & Kaliontzopoulou, 2020), Morpho (version 2.7; Schlager, 2017), rgl (version 0.100.47; Adler & Murdoch, 2020), RRPP (version 0.5.2; Collyer & Adams, 2018, 2020), shapes (version 1.2.5; Dryden, 2019), and vegan (version 2.5-4; Oksanen et al., 2019). The landmark sets were imported into R, and the missing (semi)landmarks were estimated (*estimate.missing*) to allow for the incomplete landmark sets being analyzed together with the others. The following five subsets of (semi)landmarks were defined: all fixed landmarks (including start and endpoints of curves) to analyze the overall geometry of the pectoral girdle, and all landmarks and semilandmarks of a given pectoral girdle bone to allow for a more detailed shape comparison. Species lacking a clavicula were excluded from the subset consisting of (semi)landmarks on the clavicula. The following steps were performed for the full landmark sets and for each subset separately. All landmark sets of a given species were superimposed using a Generalized Procrustes Analysis (GPA; *gpagen,* if applicable, including sliding of semilandmarks to minimize bending energy), rescaled to their original centroid size and the species mean shape was calculated (*mshape*). A GPA (including sliding of semilandmarks to minimize bending energy, if applicable) was performed to superimpose the species mean shapes. The resulting sets of superimposed species mean shapes will subsequently be referred to as full landmark dataset and landmarks datasets i through v, with the full dataset consisting of all landmarks and semilandmarks, landmarks datasets i denoting the set of fixed landmarks and ii-v denoting the sets comprising all landmarks and semilandmarks, respectively, on the scapula, coracoid, cleithrum, or clavicula.

Statistical analyses and visualization

The full landmark dataset was used to assess the modularity (sensu Schlosser, 2002) within the pectoral girdle in a phylogenetic context by calculating the covariance ratio (*phylo.modularity*; Adams, 2016); modules are constituted by highly correlated subsets of traits (here landmark coordinates), whereas the covariation between such modules is relatively weak. The statistical significance was assessed by 1,000 permutations.

The following analyses were performed for each set of superimposed species mean shapes (landmark datasets i-v) separately. The phylogenetic signal in the landmark data was

assessed using a multivariate version of the K-statistic with the statistical significance being determined by 1,000 random permutations (*physignal*; Adams, 2014). As there was a statistically significant phylogenetic signal in all landmark datasets, separate phylogenetic MANOVAs (pMANOVAs; using residual randomization and type-II sums of squares; *procD.pgls*) were performed to test for significant differences between the mean shapes of locomotor groups. Potential effects of specimen size on shape were accounted for by incorporating the log-transformed centroid size and its interaction with mode of locomotion in the pMANOVAs. If there were statistically significant differences, pairwise comparisons of the mean shape between locomotor groups were performed while accounting for size (*pairwise*; null model: *coords ~ logCS*, where *coords* denotes one of landmarks sets i-v and *logCS* the log-transformed centroid size). Statistical significance was assessed by 1,000 permutations in pMANOVAs and pairwise comparisons; *p*-values below 0.05 were considered significant in all tests.

Principal component analyses (PCAs; *gm.prcomp*) were separately performed for landmark datasets i-v to visualize the distribution of species mean shapes in morphospace (*plot*, *shapeHulls*). For the dataset of fixed landmarks only (i), all individual landmark configurations belonging to a given species were transformed as their respective mean shape had been transformed during GPA and PCA (details in Engelkes et al., 2019); the transformed landmark configurations were plotted along with their means. The number of significant principal components was determined using the broken-stick model (Macarthur, 1957; *evplot* function published with Borcard, Gillet, & Legendre, 2011). Surfaces and landmark configurations were rendered in MODO (version 10.1v2; The Foundry).

Muscle moment arms

Musculoskeletal models were created for representative specimens of selected species (Figures 3 and 4) that appeared interesting based on their position in the morphospaces of the overall pectoral girdle shape (landmark dataset i) and of the coracoid shape (landmark dataset iii). The shape analyses suggested that most locomotor-mode-related shape differences occurred in the ventral part of the pectoral girdle. Therefore, the effects of the shape of the ventral pectoral girdle part (i.e., clavicula and coracoid) on the moment arms of muscles across the shoulder joint were assessed. Models were created in OpenSim (version 3.3; Delp et al., 2007) using simplified (inner structures removed, all holes in the surface closed, polygon count reduction and smoothing) surfaces of the respective specimens.

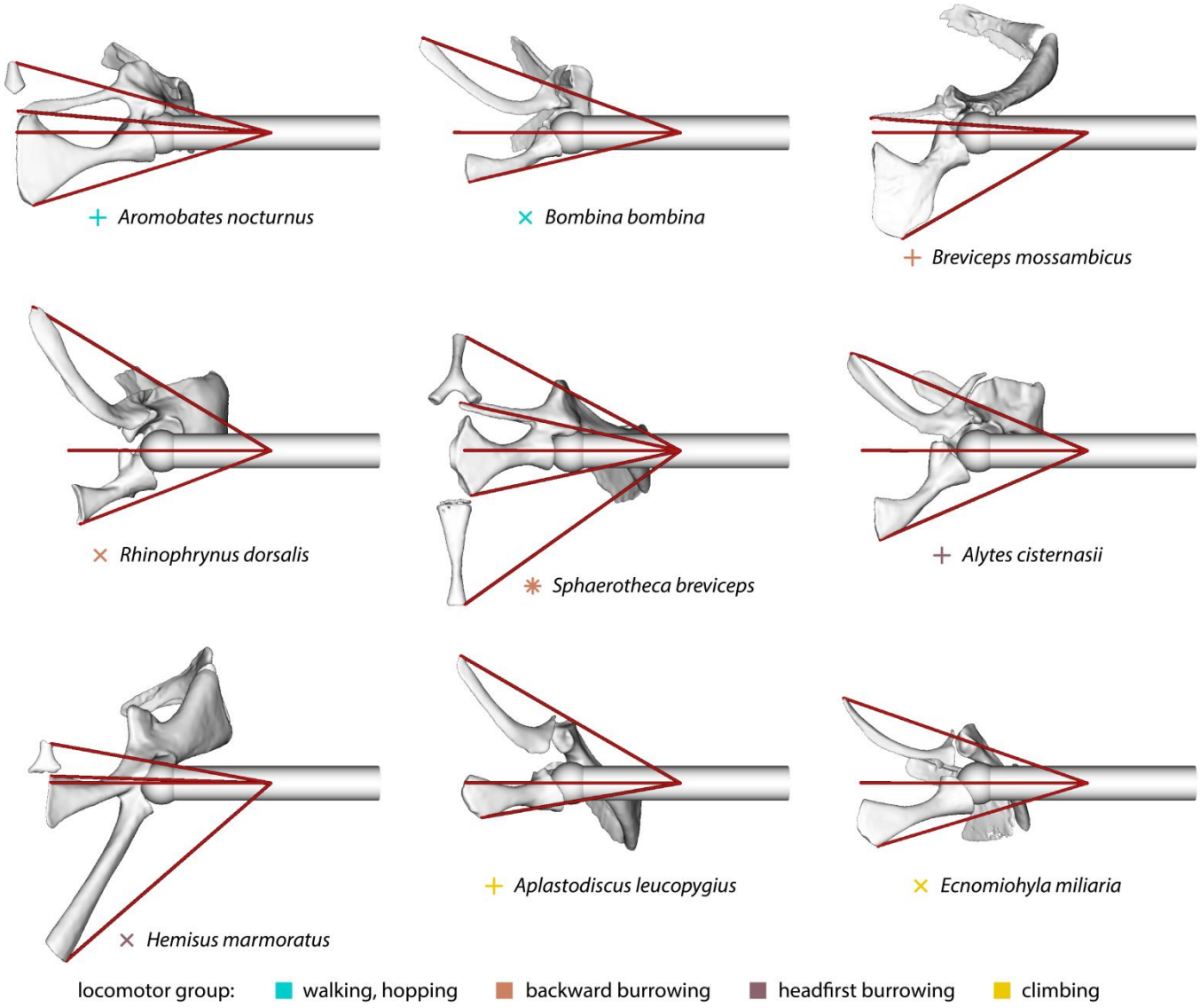


FIGURE 3. Musculoskeletal models of left-side pectoral girdle bones of selected anuran specimens. Ventral views, anterior to the top, medial to the left. Warping objects not shown. Symbols and colors as in Figure 4

Both (left and right) landmark configurations of a given specimen were combined to one configuration. This configuration was used to transform the corresponding surfaces of the left-side pectoral girdle bones and, if applicable, the bony part of the sternum or episternum to a common size and comparable orientation (R, MeshLab, MODO). The origin of the coordinate system was located in the shoulder joint cavity, the *y*-*z*-plane was parallel to the sagittal plane with the *z*-axis being approximately parallel to the long axis of the specimen, and the line connecting the anteromedial tip of the clavicula to the posteromedial tip of the coracoid was parallel to the *x*-*z*-plane. All commonly scaled and orientated surfaces were equipped with the same simplified humerus in order to exclude any potential effects of the humerus shape on muscle moment arms. The shoulder joint was defined with two axes of

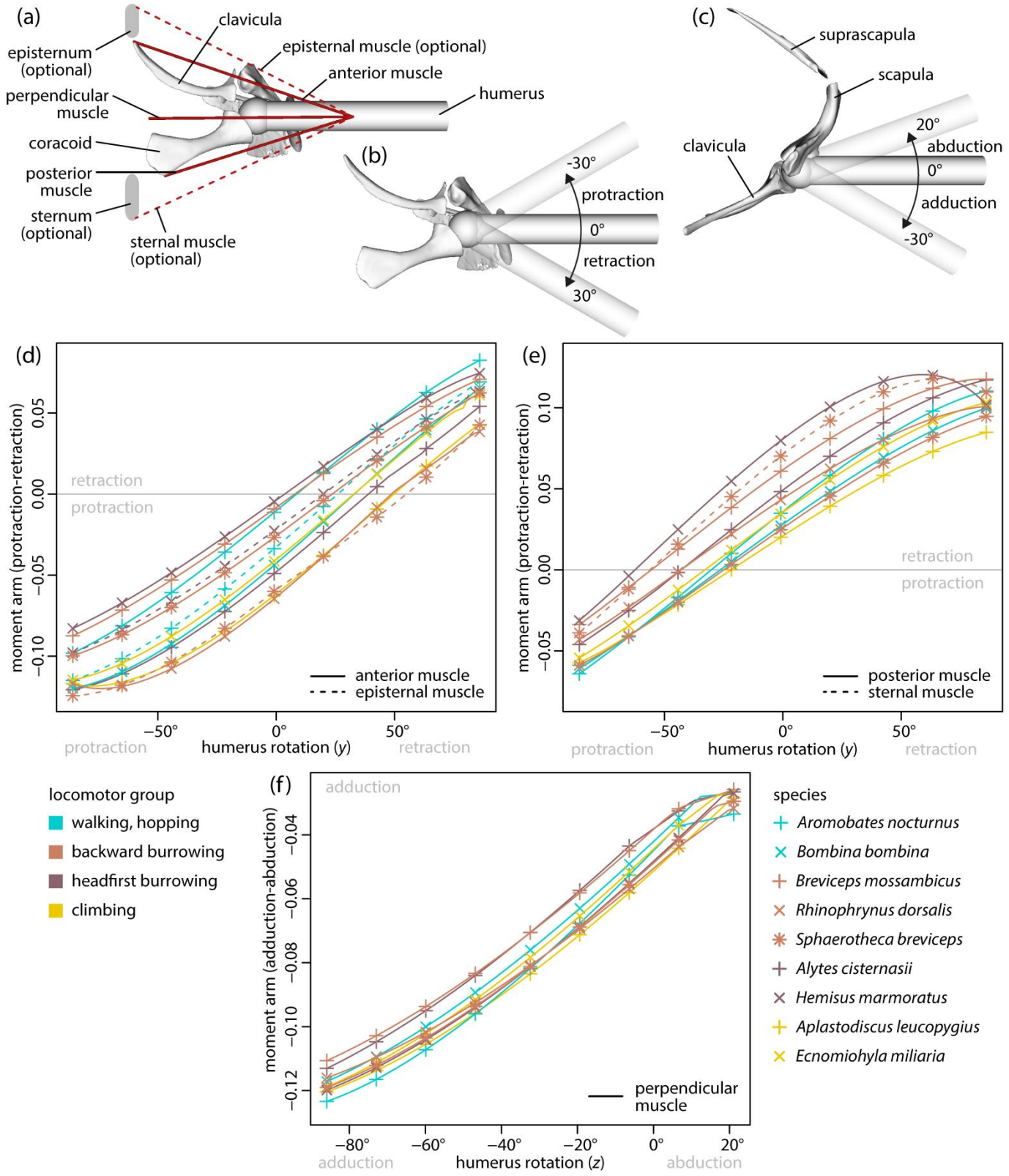


FIGURE 4. Hypothetical muscles analyzed in musculoskeletal models and respective muscle moment arms. Warping objects not shown. (a) Ventral view of musculoskeletal model of *Ecnomiohyla miliaria* with added structures that are optionally present in some specimens. (b) View of (a) without muscles to illustrate humerus protraction and retraction. (c) Anterior view of (b) to illustrate humerus adduction and abduction. (d) Moment arms of anterior and episternal muscles with regard to protraction and retraction. (e) Moment arms of posterior and sternal muscles with regard to protraction and retraction. (f) Moment arms of perpendicular muscle with regard to adduction and abduction

rotation: one allowing adduction and abduction, one allowing protraction and retraction. The humerus being aligned with the *x*-axis (perpendicular to the sagittal [*y*-*z*] plane) was used as reference position for angular measurements.

Previous studies (e.g., Bigalke, 1927; Gaupp, 1896; Ritland, 1955) showed that different muscles originated along the midline of the ventral side of the pectoral girdle and inserted onto the humerus. Those muscles were reduced to three hypothetical muscles that were included in each model (Figure 4a): one muscle ("anterior") originating from the anteromedial tip of the clavicula, one ("perpendicular") being perpendicular to the long axis of the specimen, and one ("posterior") originating from the posteromedial tip of the coracoid. If an episternum or sternum was present and contained a pars ossea (senus Gaupp, 1896), additional muscles ("episternal", "sternal", respectively) originating from, respectively, the anterior or posterior tip of the bone were included, too. All muscles were defined to insert in a common point at the humerus. Warping objects were configured as needed to prevent muscle pathways from intersecting with skeletal elements; the potential effects of soft tissues in shaping muscle pathways were neglected. The moment arms of the perpendicular muscle was assessed with regard to adduction and abduction, the moment arms of all other muscles were determined with regard to protraction and retraction (Figure 4b, c).

Finite element analysis of coracoids

The species close to the extreme ends of the first two principal components and a species close to the mean shape in the landmark dataset of species mean coracoid shapes (iii) were chosen to assess the effects of different loading conditions by using finite element analyses. The coracoid surfaces were extracted from the corresponding surfaces used for musculoskeletal modelling. As a consequence of this approach, all coracoids were scaled and orientated in a way that reflected the actual conditions in the specimens and they were modelled as solid structures. Neglecting inner structures was expected to have a minor effect base on the observations of Mielke and Nyakatura (2019). The coracoid in *Hemisus marmoratus* was fused to other bones, those bones were manually removed (MODO).

Tetrahedral meshes were generated and the models were set up in FEBio Studio (version 1.0.0; Maas, Ellis, Ateshian, & Weiss, 2012). Bone was modeled as an isotropic elastic material with a Young's modulus of 10 GPa and a Poisson's ratio of 0.35 as these values lay within the previously reported ranges for vertebrate bones (e.g., Currey, 1984; Hudson, Bennett, & Franklin, 2004). Five different loading scenarios were deduced from supposed functions of the coracoid (Table 2; Figure 5a). The applied loads were scaled by the

TABLE 2. Loading scenarios applied to selected coracoids

Scenario	Fixed in space	Force	Purpose
I	Medial surface (interface to epicoracoid cartilage)	Compressive load along the long axis (line connecting the mean point of the anteromedial and posteromedial tips of the coracoid with the center of rotation of the shoulder joint), applied to a part of the glenoidal surface	Reference condition, as we expected this to reflect the optimal loading direction
II	Medial surface	Compressive load, perpendicular to the sagittal plane, applied to a part of the glenoidal surface	Simulation of medially directed force components, that occur during landing (Emerson, 1983) or burrowing (Emerson, 1976)
III	Part of glenoidal surface	Load (tension) along the trajectory of the hypothetical posterior muscle (musculoskeletal model in reference position), applied to a small area on the posteromedial part of the coracoid	Simulation of loading due to muscles originating in this area
IV	Part of glenoidal surface	Anteriorly directed load, parallel to the longitudinal axis of the specimen, applied to the posteromedial part of the medial surface of the coracoid	Simulation of potential anteriorly directed force component that a sternum might transmit to the pectoral girdle if muscles attached to the sternum contract and thereby pull the sternum forward
V	Part of glenoidal surface	Posteriorly directed load, parallel to the longitudinal axis of the specimen, applied to the posteromedial part of the medial surface of the coracoid	Simulation of potential effect of a m. sterno-epicoracoideus or m. rectus abdominis (Emerson, 1983; Jones, 1933) that could be attached to the posteromedial tip of the epicoracoid cartilage

area on which they were applied such that equal forces were applied across all loading scenarios and specimens. Von Mises stresses were visualized in PostView (version 2.5.0; also part of FEBio suite), and the mesh-weighted arithmetic mean von Mises stresses (mwam; Marcé-Nogué, Esteban-Trivigno, Escrig, & Gil, 2016) were calculated in R.

Results

Overall shape of pectoral girdle bones

The first five principal components (PCs) of the species mean shapes of the overall geometry of the pectoral girdle bones (landmark dataset i) were significant and, respectively, represented 48.62%, 14.42%, 8.5%, 5.62%, and 5.1% of the variance in the landmark data.

The pectoral girdle shapes of swimming and climbing species, and those of swimming and backward burrowing species differed with regard to the shape differences associated with PC 1 (Figure 6) and, in the latter case, also PC 4. In addition, there was a tendency for shape differences between backward and headfirst burrowing species along PCs 1 and 4, between burrowing and nonburrowing species along PC 2, and between backward burrowing and climbing species along PC 4. Yet, all locomotor groups comprised pectoral girdle shapes that

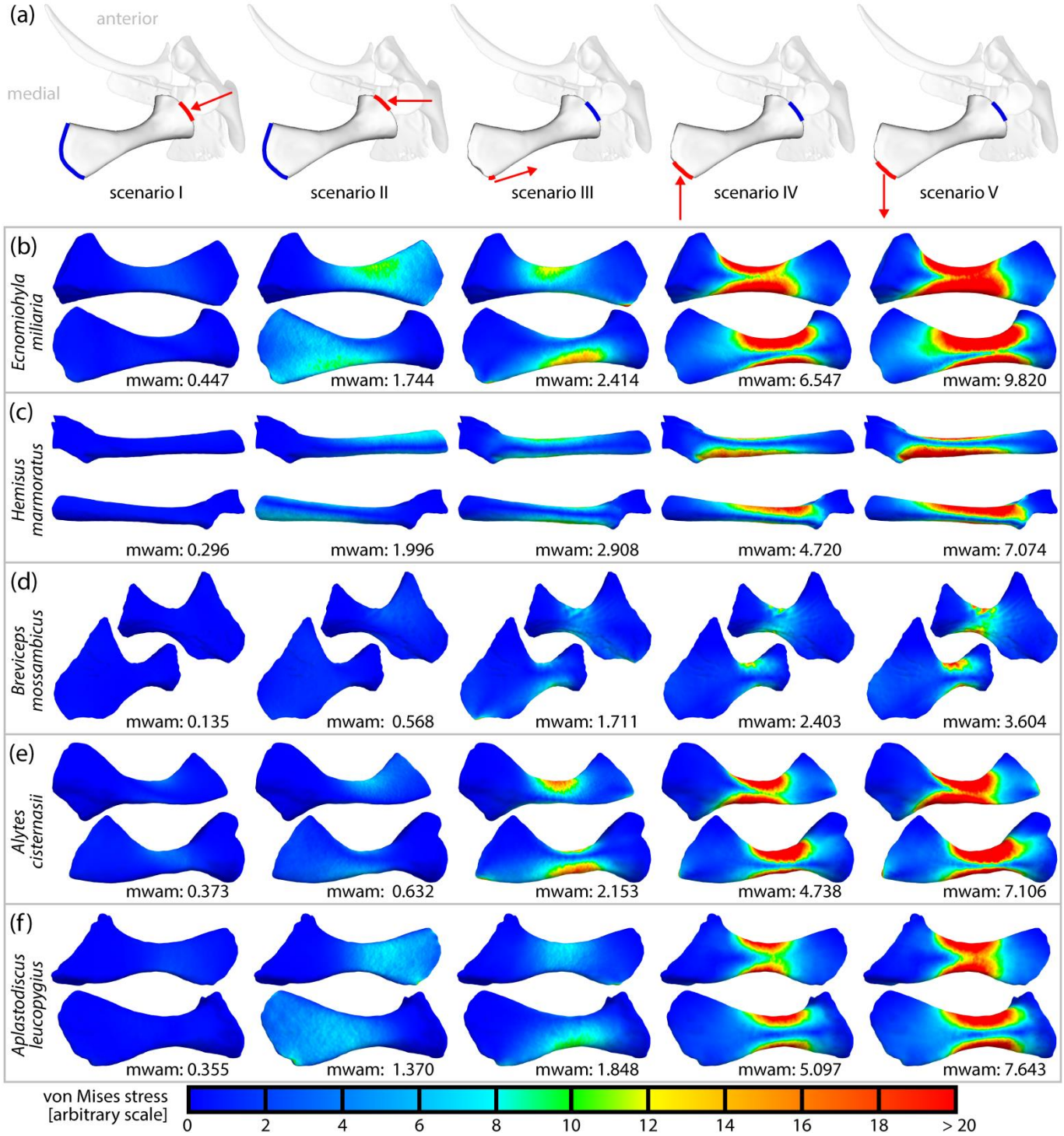


FIGURE 5. Loading scenarios and von Mises stresses in finite element analyses (FEAs) of coracoids. (a) Loading scenarios in ventral views. Blue line: surface fixed in space; red line: area of force application; red arrow: direction of applied force. (b-f) Results of FEAs. Size not comparable among species. Anterior approximately to the top; dorsal view above ventral view. mwam: mesh-weighted arithmetic mean von Mises stress

were similar to some of those observed in other groups (i.e., all locomotor groups showed some regions of overlap along PCs 1-5 in pairwise comparisons) and the species represented by more than one specimen showed some shape similarities (overlap in PC plot) with other species.

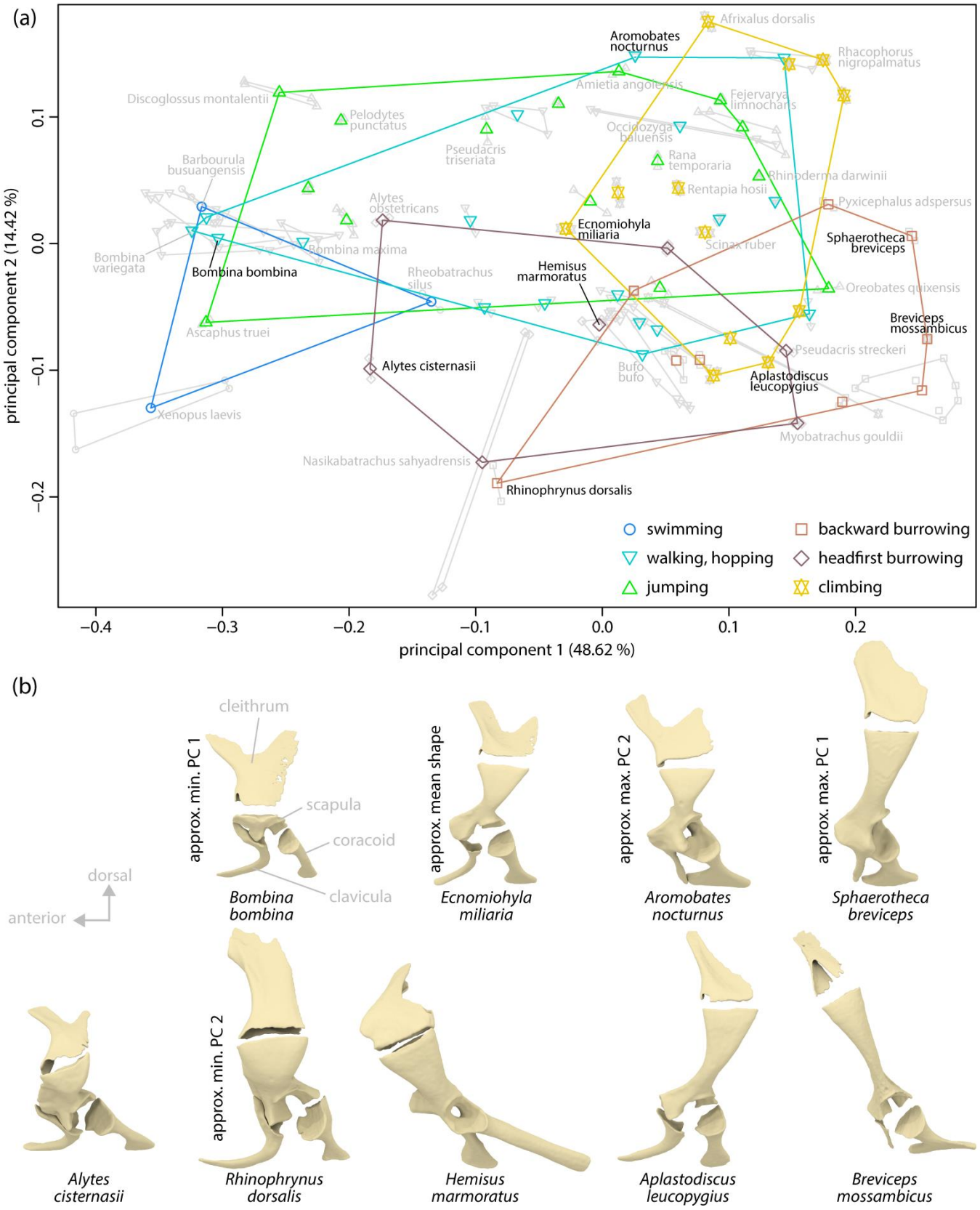


FIGURE 6. Principal component (PC) plot of overall species mean shapes of pectoral girdle bones (landmark dataset i) and surfaces of selected specimens. (a) PC plot of species mean shapes. Gray points illustrate single landmark configurations used to calculate the species mean shapes. (b) Surfaces of the left-side pectoral girdles used for musculoskeletal modelling; sternal and episternal elements omitted

TABLE 3. Results of pMANOVA of species mean shapes of pectoral girdle bones performed for fixed landmarks only (landmark dataset i)

	Df	***SS***	***MS***	R^2	***F***	***p***
Log. centroid size	1	0.0013047	0.00130473	0.05260	4.2706	0.001*
Locomotor mode	5	0.0048325	0.00096649	0.19483	3.1635	0.001*
Log. centroid size:locomotor mode	5	0.0017934	0.00035867	0.07230	1.1740	0.234
Residuals	52	0.0158866	0.00030551	0.64052		
Total	63	0.0248028				

Asterisks (*) denote statistical significance.

TABLE 4. Pairwise comparison of locomotor-group-specific species mean shapes of pectoral girdle bones (calculated from species mean shapes)

Locomotor groups compared	***p* value for overall shape**	***p* value for scapula shape**	***p* value for coracoid shape**
Swimming - walking, hopping	0.052	0.091	0.720
Swimming - jumping	0.051	0.216	0.794
Swimming - backward burrowing	0.006*	0.023*	0.265
Swimming - headfirst burrowing	0.074	0.024*	0.918
Swimming - climbing	0.018*	0.012*	0.828
Walking, hopping - jumping	0.615	0.321	0.545
Walking, hopping - backward burrowing	0.023*	0.209	0.006*
Walking, hopping - headfirst burrowing	0.003*	0.130	0.058
Walking, hopping - climbing	0.352	0.055	0.260
Jumping - backward burrowing	0.002*	0.025*	0.001*
Jumping - headfirst burrowing	0.001*	0.027*	0.179
Jumping - climbing	0.116	0.005*	0.842
Backward burrowing - headfirst burrowing	0.017*	0.947	0.001*
Backward burrowing - climbing	0.014*	0.975	0.001*
Headfirst burrowing - climbing	0.007*	0.633	0.194

Asterisks (*) denote statistical significance.

The first principal component was mainly associated with differences in the height (dorsal-ventral expansion) of the scapula relative to its width (anterior-posterior expansion) and to the length of the clavicula and coracoid, the position of the glenoid cavity relative to the dorsal margin of the scapula and the ventral midline, as well as the angles of the clavicula and coracoid relative to each other and to the ventral midline. A high scapula was generally associated with a more medially located glenoid cavity and with comparably short clavicula and coracoid; the long axes of the ventral bones lay approximately parallel to each other and rather perpendicular to the ventral midline of the specimen. If a flat scapula was present, the ventral bones were angled such that their long axes diverged medially. The clavicula was curved, and the anterior and posterior margins of the coracoid were comparably straight (inferred from exemplary pectoral girdles shown in Figure 6 as semilandmarks were not included in this dataset). The second principal component was also associated with differences in the shape and angle of the ventral bones, the length of these bones relative to the height of the scapula, and the position (in anterior-posterior direction) of the glenoid cavity relative to the dorsal margin of the scapula.

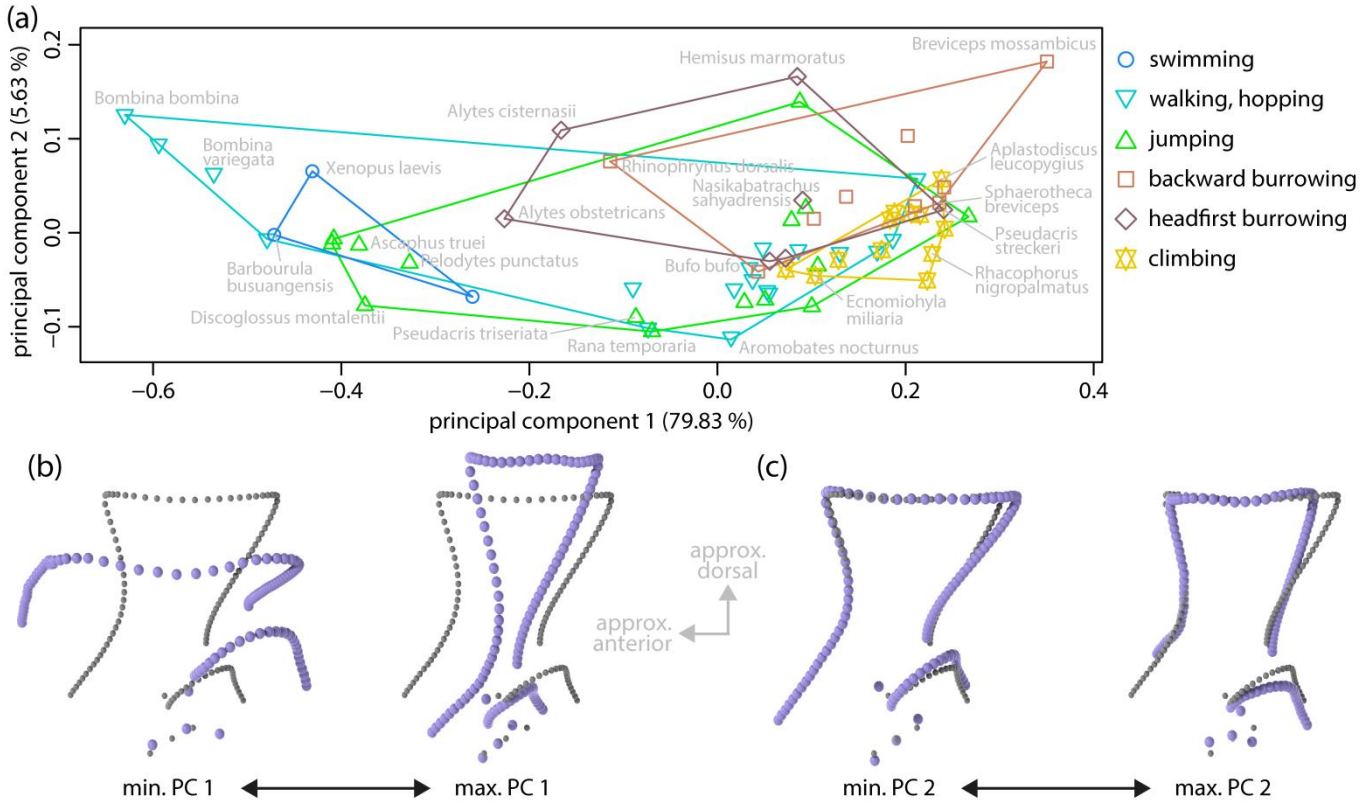


FIGURE 7. Principal component (PC) plot of species mean shapes of scapulae (landmark dataset ii) with extreme shapes along PCs. (a) PC plot of species mean shapes. (b) Extreme shapes of PC 1 in lateral view. (c) Extreme shapes of PC 2 in lateral view. Gray: mean shape; violet: extreme shape

The phylogenetic signal ($K_{mult} = 0.9595$; $p = 0.001$) and the effects of the log-transformed centroid size and locomotor mode on shape were statistically significant (Table 3). The locomotor mode ($R^2 = 0.19483$) accounted for considerably more of the shape variation than the log-transformed centroid size ($R^2 = 0.05260$). The pairwise comparison of mean shapes of locomotor groups revealed that climbing species significantly differed from swimming species, that the group of backward burrowing species significantly differed from headfirst burrowing species, and that each burrowing group significantly differed from all other locomotor groups, except for headfirst burrowers that did not differ from swimmers (Table 4). The modularity test performed on the full landmark dataset revealed significant modularity (covariance ratio: 0.8133; $p = 0.001$).

Shape of the scapula

Most shape variance (79.83%) in the species mean shapes of the scapula (landmark dataset ii) was represented by PC 1. This principal component was the only significant component and revealed a tendency towards shape differences between non-neobatrachian and neobatrachian

TABLE 5. Results of pMANOVA of species mean shapes of scapulae performed for respective fixed landmarks and semilandmarks (landmark dataset ii)

	Df	*SS*	*MS*	R^2	*F*	*p*
Log. centroid size	1	0.0010460	0.00104602	0.04950	3.7193	0.016*
Locomotor mode	5	0.0037336	0.00074672	0.17668	2.6551	0.003*
Log. centroid size:locomotor mode	5	0.0009149	0.00018298	0.04329	0.6506	0.872
Residuals	52	0.0146244	0.00028124	0.69203		
Total	63	0.0211326				

Asterisks (*) denote statistical significance.

TABLE 6. Results of pMANOVA of species mean shapes of coracoids performed for respective fixed landmarks and semilandmarks (landmark dataset iii)

	Df	*SS*	*MS*	R^2	*F*	*p*
Log. centroid size	1	0.0003308	0.00033077	0.02087	1.6416	0.129
Locomotor mode	5	0.0040459	0.00080918	0.25533	4.0159	0.001*
Log. centroid size:locomotor mode	5	0.0006368	0.00012736	0.04019	0.6321	0.894
Residuals	52	0.0104776	0.00020149	0.66123		
Total	63	0.0158455				

Asterisks (*) denote statistical significance.

anurans (Figure 1; both groups roughly separated along PC 1 in Figure 7). It was associated with differences in the height relative to the width of the scapula, and with the curvature of the anterior margin. A high scapula was associated with a concavely shaped anterior margin, whereas the corresponding structure of a low scapula was rather convex. PC 2 represented 5.63% of the variance and, despite its insignificance, was mainly associated with differences in the torsion of the scapula around its long (dorsoventral) axis, the length of the dorsal margin relative to the ventral expansion, and the angle of the dorsal margin of the glenoid cavity relative to the horizontal plane. The scapula shape of swimming species differed from the shape of burrowing and climbing species along PC 1, and there was a tendency towards shape differences between burrowing and climbing species along PCs 1 and 2.

There was a strong and significant phylogenetic signal (K_{mult} = 1.6003; p = 0.001) in the species mean shapes of the scapula. The effects of the log-transformed centroid size (R^2 = 0.04950) and the locomotor mode (R^2 = 0.17668) were statistically significant with the latter clearly exceeding the former (Table 5). The pairwise comparison of the mean shapes of locomotor groups (Table 4) revealed that jumping and swimming species significantly differed from burrowing and climbing species.

Shape of the coracoid

The first three PCs of the species mean shapes of the coracoid (landmark dataset iii) were significant and, respectively, represented 59.25%, 14.81%, and 9.15% of the total variance. Headfirst burrowing and swimming species differed from backward burrowing species along

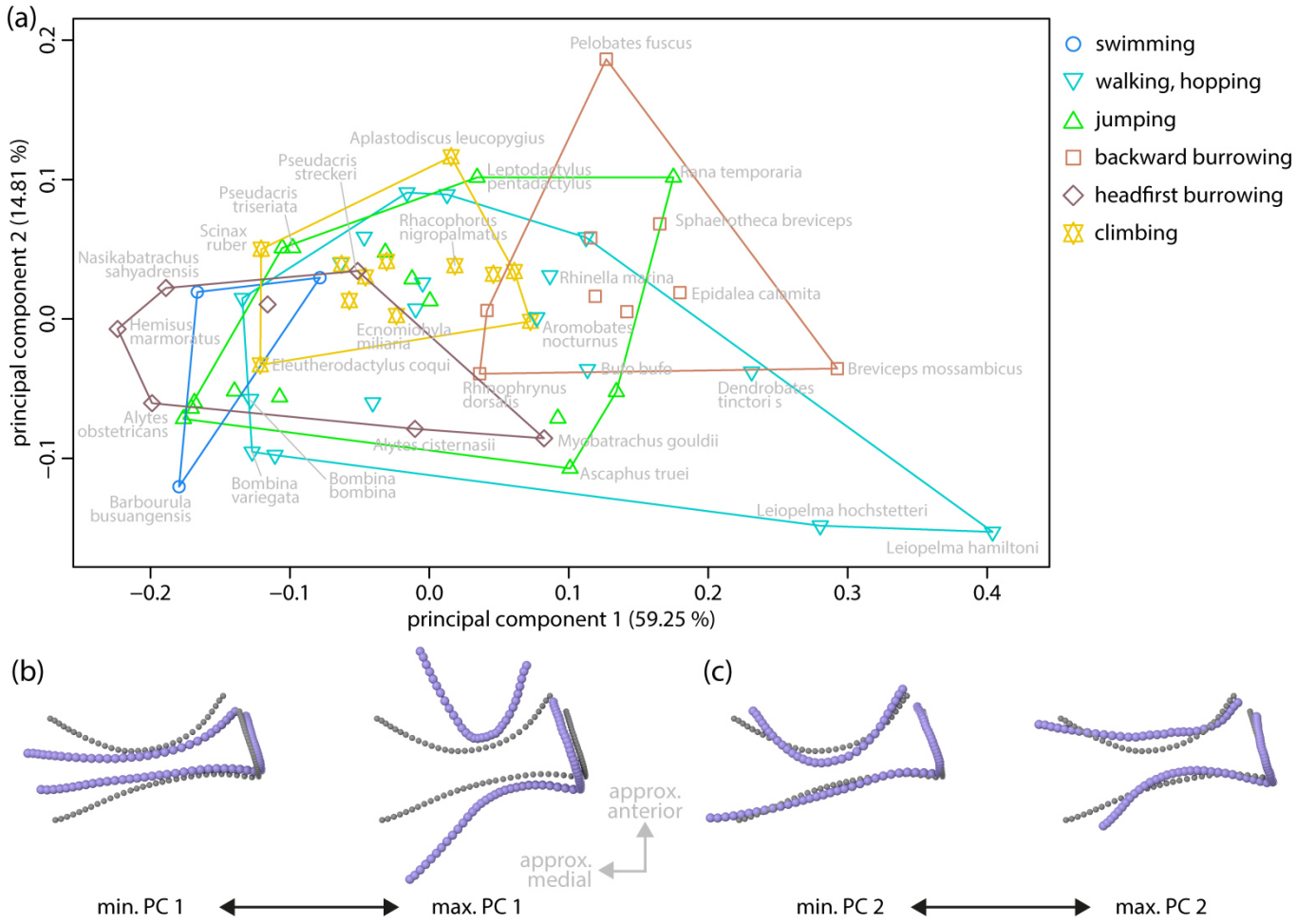


FIGURE 8. Principal component (PC) plot of species mean shapes of coracoids (landmark dataset iii) with extreme shapes along PCs. (a) PC plot of species mean shapes. (b) Extreme shapes of PC 1 in ventral view. (c) Extreme shapes of PC 2 in ventral view. Gray: mean shape; violet: extreme shape

the first two PCs. There was no specific pattern with regard to group-related shape differences along PC 3. The coracoid shapes mainly differed in their length (long axis, approx. medial-lateral expansion) relative to their width (anterior-posterior expansion) in combination with different degrees of curvature of the anterior and posterior margin (Figure 8). These shape differences were associated with PC 1. The shape variation along PC 2 mainly represented differences in the curvature of the long axis in the anterior-posterior direction in combination with differences in the curvature of the anterior and posterior margins.

The phylogenetic signal in the species mean coracoid shape was significant ($K_{mult} = 0.6355$; $p = 0.001$), yet small compared to the phylogenetic signal in the overall pectoral girdle shape and the shape of the scapula. The effect of locomotor mode on coracoid shape was significant ($R^2 = 0.25533$; Table 6) and the pairwise comparison of locomotor group mean shapes (Table 4) showed that backward burrowing species significantly differed from all other locomotor groups, except for swimming species.

Shapes of the cleithrum and clavicula

There were significant phylogenetic signals in the species mean shapes of the cleithrum (landmark dataset iv; $K_{mult} = 0.5113$; $p = 0.001$) and the clavicula (landmark dataset v; $K_{mult} = 0.8752$; $p = 0.001$). The pMANOVAs revealed no significant effects of locomotor group or log-transformed centroid size and the principal components showed no clear pattern of separation of locomotor groups for any of the two landmark datasets (iv, v), although the first three (iv) or four (v) PCs were significant.

Muscle moment arms

The moment arms (Figure 4d-f) of the hypothetical muscles showed that the action of the muscles depended on the position of the humerus. The range of humerus positions (protraction-retraction) in which the posterior muscles contributed to retraction generally was largest in burrowing species. The moment arms of this muscle for retraction also were generally larger in burrowing species. The one exception to these observations was the backward burrowing species *Sphaerotheca breviceps* which showed comparably small moment arms for the posterior muscle during retraction and had a relatively small range of humerus positions in which this muscle contributed to retraction. The sternal muscle in this species, however, showed similar properties as the posterior muscles in the other burrowing species.

The properties of the other muscles showed no clear pattern of association with locomotor groups. Similar to the sternal muscle, the existence of a episternal muscle (if present) increased the moment arm for humerus protraction and widened the range of humerus positions for which the muscles contributed to protraction if compared to the anterior muscle of the respective species.

Finite element analysis of coracoids

Within each species, lowest mesh-weighted arithmetic mean (mwam) von Mises stresses were observed if the coracoid was loaded along its long axis (scenario I; Figure 5). Highest stresses occurred if the posteromedial surface of the coracoid was pulled backward to simulate the potential effect of a m. sterno-epicoracoideus or m. rectus abdominis (scenario V), and second-highest stresses occurred if the same region was pushed forward to simulate potential forces transmitted by a sternum (scenario IV). Across species, the coracoid of *Breviceps mossambicus* experienced lowest mwam von Mises stresses under all loading scenarios. The

coracoid of *Hemisus marmoratus* experienced highest mwam von Mises stresses under loading through the shoulder joint in lateromedial direction (scenario II) or by the hypothetical action of the posterior muscle (scenario III).

Discussion

Our data indicate that the phylogenetic history, the size, and the locomotor behavior have significant effects on the shape of the pectoral girdle bones of anurans but the relative impact of these factors differs between bones. The most striking locomotor-behavior-related shape differences were observed between burrowing and nonburrowing species; those differences might be explained by a functional adaptation to the burrowing behavior and are possibly associated with trade-offs. The shapes of the other locomotor groups differed less or even not at all and most groups showed remarkable within-group shape diversity. Similarly shaped pectoral girdles provide the anatomical base for different locomotor behaviors, which indicates that the processes of many-to-one mapping (i.e., different morphologies can result in the same functional performance which might lead to a partial decoupling of morphological characters and function; Wainwright, Alfaro, Bolnick, & Hulsey, 2005) has acted during the evolution of the anuran pectoral girdle.

Modularity and phylogenetic signal

The observed differences in the relative impact of the considered factors (phylogeny, size, locomotion) on the shape of the distinct pectoral girdle bones might indicate some modularity within the pectoral girdle of anurans. This is supported by the statistical significance of the modularity test, although the result of this test should be interpreted with caution, as the test was performed on fixed landmarks and semilandmarks (Cardini, 2019).

At least some anatomical traits of anurans are influenced by their phylogenetic history; among these traits are the absolute and relative length of the hind limbs (Gomes, Rezende, Grizante, & Navas, 2009), the relative length of the tibiofibula and femur, their ratio, and the snout-vent length (Enriquez-Urzelai et al., 2015), the relative length of the foreleg (Vidal-García, Byrne, Roberts, & Keogh, 2014), and several other external body dimensions (Sherratt, Vidal-García, Anstis, & Keogh, 2017). Our results are in line with these previous studies as the species mean shapes of the entire pectoral girdle and of its distinct bones showed a significant phylogenetic signal. There were differences, however, in the relative strength of the phylogenetic effect on the shapes of the single bones as indicated by different

values of K_{mult}. The species mean shapes of the scapulae resembled each other more than expected under a Brownian motion model ($K_{mult} > 1$), which implies that the phylogenetic history is the major factor in the evolution of the scapula shape. This is also supported by the observed differences in the shapes of the scapulae of non-neobatrachian and neobatrachian species (Figures 1 and 7). The effects of size and locomotor mode, although statistically significant, seem to influence the scapula shape to a minor extent. In contrast, the observed phylogenetic signal in the species mean coracoid shape was comparably small and below the expectation under Brownian motion ($K_{mult} < 1$). This indicates that other factors (i.e., locomotion) besides phylogeny influence the evolution of the coracoid shape.

Among the factors considered herein, the phylogenetic relation seems to be the only factor to determine the shapes of the cleithrum and clavicula as the statistical analyses were insignificant for the factors size and locomotor group. But this might be an artefact caused by the GPA or pMANOVA, as the shape of each of these bones was analyzed using one curve of more or less colinear semilandmarks only. There might be an association of the shape of these bones with size or locomotion that was not detected by our analyses.

These observations allow the hypothesis that the evolution of the shape of the distinct pectoral girdle bones is driven by different primary factors, although they are part of the same complex. If so, this could indicate differences in the functional importance of these bones.

Adaptation of pectoral girdle shape to burrowing behavior

The most striking differences in the pectoral girdle shape were observed between burrowing and nonburrowing species (Figures 6 and 8; Table 4), which is in general accordance with previous studies that reported burrowing behavior to be associated with modifications of various anatomical structures (summarized in the introduction). The mean pectoral girdle shapes of backward and headfirst burrowing species significantly differ from one another and from other locomotor groups in one or more aspects (Table 4), indicating that the pectoral girdle bones of burrowing frogs may be specifically adapted to burrowing behavior. In particular, increased moment arms of the humerus retractor muscles (herein modeled as the posterior muscle) and widened ranges of humerus positions, for which this muscles acts as a retractor, were observed for most burrowing species if compared to nonburrowing species by musculoskeletal modelling (Figure 4e). This might be explained by specific biomechanical requirements linked to burrowing.

Emerson (1976) observed that specimens of the headfirst burrowing species *Hemisus marmoratus* moved the soil by forelimb retraction and that this motion was accompanied by a

lateral force component. She assumed the enlarged retractor muscles and the elongated, posteriorly angled coracoids found in this species to be adaptations to the headfirst burrowing behavior. Our results indicate additional effects of the shape and orientation of the coracoid: The specific configuration of the coracoid shifted the origin of the posterior muscle backwards and thereby increased its moment arm across the shoulder joint, that is, its effectiveness (Sherman, Seth, & Delp, 2013) in humerus retraction if compared to other species (Figure 4). In addition, the posterior muscle functioned as a humerus retractor in a more anterior humerus position. Both these effects seem to be advantageous for headfirst burrowing and, thus, likely are adaptations to the burrowing behavior of *H. marmoratus*.

The finite element analyses revealed that the coracoid of *H. marmoratus* experienced comparable high mesh-weighted arithmetic mean von Mises stresses if loading by the posterior muscle (scenario III) or by mediolateral compression (scenario II) was simulated (Figure 5c). This is somewhat surprising as both these loading scenarios seem ecologically relevant: The posterior muscle simulated the forces produced by the humerus retractor muscles, and there is a lateral force loading the pectoral girdle during headfirst digging (compare Emerson, 1976). The comparably high von Mises stress might be a trade-off for the enlarged muscle moment arms across the shoulder joint caused by the elongation and specific orientation of the coracoid. It should be noted that the force of the posterior muscle was simulated for the humerus being orientated perpendicular to the sagittal plane; the observations of Emerson (1976) indicate that highest digging forces might occur in a more anterior humerus position. If so, the peak force imposed by the posterior muscle would be more aligned with the long axis of the coracoid, which in turn could result in smaller mean von Mises stress (also compare scenario I).

With regard to the pectoral girdle resisting to medially directed compression, it is noteworthy that the clavicula in *H. marmoratus* is angled rather perpendicular to the ventral midline, more robust, and enlarged medially (Figure 6b; also see Braus, 1919; Emerson, 1976). This shape and orientation somewhat resemble the configuration of the coracoid in some other species, and we hypothesize that, in *H. marmoratus*, the clavicula replaces the coracoid, for example, in transmitting and dissipating medially directed compressive forces through the shoulder joint. If this was true and the clavicula resisted most of the forces imposed by a medially directed compression, the mediolateral bending of the coracoid would be considerably reduced, which in turn would have led to smaller von Mises stresses in the coracoid. Such an effect was not observed in our simulations as we artificially removed the

clavicula and the scapula, but the fusion of these two bones to the coracoid (Figure 6b) might be an indicator for their interaction in force transmission.

The specific clavicula configuration observed in *H. marmoratus* results in a small moment arm for the anterior muscle (Figure 4d) with regard to humerus protraction. Such small moment arms with regard to humerus protraction should be a disadvantage for headfirst burrowing as the retracted humerus needs to be moved forward for a new digging cycle. The bony episterum in *H. marmoratus* might have evolved to compensate for this disadvantage by expanding the area for muscle attachment anteriorly, which in turn leads to a larger moment arm across the shoulder joint (see episternal muscle in Figure 4d; also compare Trueb, 1973).

Large moment arms for the humerus retractor muscles seem to be a requirement for backward burrowing, too (compare Figure 4e), but the reason for this is not as obvious as for headfirst burrowing. To our knowledge, no detailed description of the function of the forelimbs (i.e., the forces acting on them) during backward burrowing does exist. The forelimbs are species-specifically either used to stabilize the body (Emerson ,1976; Sanders & Davies, 1983) or to turn the body in the excavated hole (Sanders & Davies, 1983) during backward burrowing. Considering these functions, it might be hypothesized that the humerus retractor muscles mainly act to stabilize the shoulder joint while digging with the hind limbs, but this needs to be investigated in future studies.

It is remarkable that the coracoid of *Breviceps mossambicus* experienced lowest von Mises stresses in the finite element analyses (Figure 5d). Among the simulated loading scenarios, the resistance to lateral compression (scenario II) and to forces imposed by the humerus retractor muscles (scenario III) seem to be the most ecologically relevant, as backward digging is associated with a lateral force component (Emerson, 1976) and the retractor muscles likely are active during digging. The specific coracoid shape may thus be an adaptation to the backward burrowing behavior in *B. mossambicus* and comes at the cost of a small moment arm of the posterior muscle with regard to humerus retraction (Figure 4e). Analogous to the episternum in *H. marmoratus* (and other species), the pars ossea of the sternum in *B. mossambicus* might have evolved to compensate for this presumably disadvantageous moment arm (also compare Trueb, 1973). Cartilaginous episternal or sternal elements, as described for various species (e.g., Braus, 1919; Fürbringer, 1873; Trueb, 1973), were not considered herein. Yet, they might have a similar advantageous effect on muscle moment arms across the shoulder joint and should be included in future studies.

Two further observations support the hypothesis that the pectoral girdles of different species are adapted to their burrowing: *Alytes cisternasii* has been reported to be the faster and

more efficient headfirst burrower if compared to the also headfirst burrowing *A. obstetricans* (Brown & Crespo, 2000). This coincides with the pectoral girdle shape of *A. obstetricans* being within the range of walking, hopping and jumping species, whereas the shape of *A. cisternasii* more resembles that of other headfirst burrowing species (Figure 6a). Thus, some anatomical specialization in the pectoral girdle of *A. cisternasii* might allow this species to perform better in burrowing. Despite the significant phylogenetic signal, the shape differences in the pectoral girdles of the jumping species *Pseudacris triseriata* and the headfirst burrowing species *P. streckeri* are comparably large with the latter more closely resembling the shape of other burrowing species (Figure 6a).

Walking, hopping and jumping

In contrast to previous studies (see introduction for a summary), our analyses indicate that there is no specific pectoral girdle shape associated with either of these locomotor modes (Table 4) and, in particular, both locomotor groups do not differ in their mean pectoral girdle shape. Instead, walking, hopping and jumping species display a remarkable within-group shape diversity in the pectoral girdle bones and their orientation to one another (Figures 6-8). It appears that differently shaped pectoral girdles are equally suited to fulfill the biomechanical requirements of jumping or walking, hopping.

Swimming and climbing

Swimming species significantly differed from headfirst burrowing species, as well as climbing from jumping species in the mean shapes of the scapulae only. These differences, although observed in the context of locomotor behavior, could be caused by the phylogenetic structure of the respective locomotor groups: The group of swimming species consisted of mostly non-neobatrachians whereas the group of headfirst burrowing species consisted of non-neobatrachians and neobatrachians (Figure 1). Given the strong phylogenetic signal in the scapula shape, this unequal phylogenetic pattern in locomotor group composition alone might have separated both groups in morphospace and there might be no true shape difference caused by differences in the locomotor behavior (also see the discussion of group dispersion across the phylogeny in Adams & Collyer, 2018).

The potential lack of a specific pectoral girdle shape within aquatic species might be explained by the fact that most anurans are good swimmers and likely have pectoral girdles that allow for an efficient aquatic locomotion. If so, the pectoral girdle shape of purely aquatic species would not differ much from nonaquatic species. An additional explanation for the

nonspecific pectoral girdle shape of swimming anurans might be that the forelimbs are involved in swimming to only a minor extent (Abourachid & Green, 1999; Gillis & Biewener, 2000) and thus likely impose rather unspecific biomechanical requirements on the pectoral girdle. In addition, the effect of gravity is reduced in water (Zug, 1971) which would result in, among other things, minor forces acting on the pectoral girdle. Instead of being optimized for a high locomotor performance, the pectoral girdle of aquatic anuran species might be adapted to other ecologically relevant tasks like suction feeding (Cundall, Fernandez, & Irish, 2017). The morphological adaptation to swimming might have primarily occurred in other anatomical traits (Gómez & Lires, 2019; Laurent, 1964; Moen, 2019; Videler & Jorna, 1985).

Following the lines of argumentation above, it might be possible that there is no locomotor-behavior-related shape difference between jumping and climbing species, as the latter group consisted of neobatrachian species only, whereas the former additionally contained non-neobatrachians. It is noteworthy that climbing evolved several times independently within the Neobatrachia only (Reilly & Jorgensen, 2011). Considering the phylogenetic distribution of arboreality, some specific anatomical novelties might have evolved in the last common ancestor of neobatrachian anurans and might have been necessary for the evolution of climbing behavior. The development of a fibrous epidermis with modified mucus glands on the finger and toe pads seems a promising candidate for such a novelty, as these specifications are not present in the non-neobatrachian species *Ascaphus truei*, *Alytes obstetricans*, and *Scaphiopus holbrookii* (Noble & Jaeckle, 1928). In addition, those glands evolved before arboreality in certain anuran linages and were suggested to lead to climbing ability if combined with enlarged toe pads (Noble & Jaeckle, 1928). The lack of such a novelty might have constrained non-neobatrachians from developing climbing behavior. Given that neobatrachians have comparably high scapulae and that climbing has evolved within neobatrachians only, these specific shapes seem to be associated with climbing although the true reason for the association likely is phylogenetic relatedness. All this is speculative at this stage and requires further investigation.

Many-to-one mapping and trade-offs

The locomotor groups in our study showed a remarkable within-group pectoral gridle shape diversity (Figures 6-8). Differently shaped pectoral girdles within a given locomotor group, thus, provide the anatomical base for similar locomotor behavior. This phenomenon of different forms allowing similar functions is known as many-to-one mapping (Wainwright et al., 2005) and, although not named as such, has previously been indicated for the anuran

pectoral girdle. Arciferal and firmisternal pectoral girdles showed no considerable differences in patterns of deformation if compressively loaded through the shoulder joint (Emerson, 1984) and should thus be equally suited to accomplish tasks that require the resistance to lateral forces. Both girdle types, however, differ in the mechanism of how these forces are dissipated (Emerson, 1983; also see Figure 5). One additional example of many-to-one mapping has been observed in our study: Similar moment arms of the posterior muscle are produced by different pectoral girdle shapes in burrowing species (Figure 4). The coracoids in the respective girdles presumably accomplish different functions, namely either shifting the attachment area of the posterior muscle posteriorly or resisting mediolateral forces.

We observed few, if any, significant shape differences between swimming, jumping, climbing and walking, hopping species, and large regions of overlap of locomotor groups in morphospace (Figures 6-8). Similarly shaped pectoral girdles, thus, provide the anatomical base for different locomotor behaviors. This might be associated with trade-offs imposed by conflicting biomechanical demands (Herrel, van Damme, Vanhooydonck, Zaaf, & Aerts, 2000). On the other hand, many-to-one mapping is thought to allow for the simultaneous optimization of multiple biomechanical properties (Wainwright, 2007; Wainwright, Alfaro, Bolnick, & Hulsey, 2005), so that a given pectoral girdle shape might be equally adapted to several locomotor behaviors without functional trade-offs.

Both, many-to-one mapping and trade-offs, might have occurred during the evolution of the morphological diversity in anurans. For example, Moen (2019) observed many-to-one mapping in the relative hind limb length and relative hind limb muscle mass onto swimming and jumping performance. Neither trade-offs nor coupled optimization between the independently evolved (Abourachid & Green, 1999; Astley, 2016) locomotor modes of swimming and jumping were observed for the hind limb anatomy of a semiaquatic frog (Nauwelaerts, Ramsay, & Aerts, 2007). Anurans with different pelvic and hind leg morphologies showed similar swimming abilities and that there was no trade-off with jumping performance (Gal & Blake, 1987). These reports indicate many-to-one mapping (but see Robovská-Havelkova et al., 2014 for a report of species with different ecologies showing different kinematic patterns of hind limb motion during swimming). A trade-off has been reported between the maximum jumping distance and the jumping endurance with larger jumping distances being accompanied by an earlier onset of fatigue (Rand, 1952; Zug, 1978, 1985). Additionally, the relatively short legs of burrowing species are thought to be a trade-off between efficient burrowing and jumping performance (Gomes et al., 2009). With regard to the anuran pectoral girdle, further studies are needed to analyze the biomechanical

properties and resulting locomotor performances in order to assess which mechanisms, many-to-one mapping, trade-offs, or both, acted during the evolution of this functional complex.

Potentially undetected adaptation of pectoral girdle shape to function

Despite our observations, there might be some functional adaptation of the pectoral girdle shape to more specific motion patterns than implied by our coarse definitions of walking, hopping, jumping, swimming, and climbing. Following Emerson (1979), we defined walking, hopping and jumping locomotion based on the maximal leap length achieved by a given species. The length of a leap is determined during the initial phase of a jump by the amount of propulsive forces generated by the hind limbs (Hirano & Rome, 1984). If active at all, the forelimbs only raise the body and control the takeoff angle and do not contribute much to force generation (Akella & Gillis, 2011; Wang et al., 2014). This means that the pectoral girdle experiences comparably low forces during the initial phase and there might be no selective pressure for a specific girdle shape or function. Different landing behaviors have evolved within hopping or jumping anurans ranging from “belly flops” that do not involve the forelimbs, to coordinated landing during which the impact forces are transmitted and dissipated by the forelimbs and the pectoral girdle (Emerson ,1983; Essner, Suffian, Bishop, & Reilly, 2010; Griep et al., 2013; Reilly et al., 2016). Likewise, various landing patterns have been observed in an arboreal frog (Bijma, Gorb, & Kleinteich, 2016) and some climbing species are capable of parachuting or gliding (Oliver, 1951; also see Appendix S1: Tables A1, A2). It seems reasonable to assume that these different landing behaviors, as well as parachuting and gliding, are associated with different force patterns that act on the pectoral girdle and require specific skeletal and muscular geometries to be dissipated, particularly as landing force can be up to three times higher than the forces generated during takeoff (Nauwelaerts & Aerts, 2006).

The forelimbs of anurans are involved in other species-specific behaviors besides locomotion as, for example, prey manipulation (Gray, O'Reilly, & Nishikawa, 1997) or wiping of the body surface (Blaylock, Ruibal, & Platt-Aloia, 1976). The shapes of the pectoral girdle bones might be functionally adapted to these specific motion patterns and, given the significant phylogenetic signal and the potential effects of many-to-one mapping, might occur on a smaller scale within closely related groups. These hypotheses were not test herein.

The literature record on anuran behavior and our definition of locomotor groups might be insufficient to fully represent the behavior of at least some species. For example, the

backward burrowing species *Rhinophrynus dorsalis* is hypothesized to be capable of headfirst burrowing (Trueb & Gans, 1983). *Aplastodiscus leucopygius* is an arboreal species (Ferreira et al., 2008; Haddad & Sawaya, 2000), but at least the males have been observed to use their heads for the construction of subterranean nests that serve for egg deposition (Haddad & Sawaya, 2000). Both these species are located within or close to the region acclaimed by the group of headfirst burrowing anurans in morphospaces (Figures 6-8). Headfirst burrowing might thus require a pectoral girdle with specific biomechanical properties (potentially realized by different morphologies) and there, thus, might be adaptations to locomotor behavior that were not detected by our approach.

Limitations and future perspectives

Most anuran species in our sample were represented by one specimen only, and shape analyses were performed on the mean shapes of species. We did not consider sexual dimorphism, although this phenomenon has been reported for the humerus in some species (Lee, 2001; Padhye, Jadhav, Sulakhe, & Dahanukar, 2015; Petrović, Vukov, & Kolarov, 2017) and some muscles originating from the pectoral girdle (Emerson, 1990; Lee, 2001; Oka, Ohtani, Satou, & Ueda, 1984). Sexual dimorphism may, thus, be expected to occur in the pectoral girdle bones, too. Nevertheless, we expect these limitations to have a minor effect on our results, as the shapes of all landmark sets of a given species lay mostly within the same respective locomotor group in morphospace or expanded the region claimed by the locomotor group toward more extreme shapes without enlarging the overlap with other locomotor groups (Figure 6). Yet, sexual dimorphism and intraspecific variability in the shape of the anuran pectoral girdle bones would be interesting topics for future studies and, if combined with behavioral and biomechanical analyses, could shed light on the functional and ecological consequences of shape differences.

Muscle moment arms were simulated using a simplified humerus with all hypothetical muscles inserting at the same point in order to assess the effects of different pectoral girdle geometries independent of other factors. As Emerson (1991) argued, the length of the humerus and the location of the muscle attachments along its length influence the resulting mechanical advantage. Thus, our analysis explored only one aspect among the factors determining the biomechanical properties of the shoulder joint. Assessing the combined effects of pectoral girdle and humerus shape, as well as the consideration of species-specific muscle configurations, could provide further insight into the functionality of this complex and explain its evolution.

Acknowledgements

We thank Sebastian Büsse and Stanislav N. Gorb of the Functional Morphology and Biomechanics Group of the Kiel University for providing access to the CT scanner. We are also grateful to the following persons and their respective museums for specimen loan: David A. Kizirian and the American Museum of Natural History, Jens Vindum and the California Academy of Sciences, Ulrich Scheidt and Konrad Kürbis of the Naturkundemuseum Erfurt, Alan Resetar and the Field Museum of Natural History, Kevin de Queiroz of the National Museum of Natural History, Mark-Oliver Rödel and Frank Tillack of the Museum für Naturkunde Berlin, and Frank Glaw and Michael Franzen of the Zoologische Staatssammlung, Munich. We further thank Kai Wörner and Stefan Thiemann (Universität Hamburg) for their help with publishing the CT volumes and the two anonymous reviewers for their valuable comments that greatly improved the manuscript. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 387723284. Additional funding came from the Wilhelm-Peters-Fonds of the Deutsche Gesellschaft für Herpetologie und Terrarienkunde e.V (DGHT).

Author contributions

Karolin Engelkes: Conceptualization (lead); Data curation (equal); Formal analysis (lead); Funding acquisition (supporting); Investigation (lead); Methodology (lead); Project administration (lead); Software (lead); Supervision (equal); Validation (lead); Visualization (lead); Writing-original draft (lead); Writing-review & editing (equal). Lena Kath: Data curation (equal); Investigation (supporting); Writing-review & editing (supporting). Thomas Kleinteich: Investigation (supporting); Resources (supporting); Writing-review & editing (supporting). Jörg Hammel: Investigation (supporting); Resources (supporting); Writing-review & editing (supporting). André Beerlink: Investigation (supporting); Resources (supporting); Writing-review & editing (supporting). Alexander Haas: Data curation (equal); Funding acquisition (lead); Supervision (equal); Writing-original draft (supporting); Writing-review & editing (equal).

Data availability

CT volumes can be downloaded from https://www.fdr.uni-hamburg.de/search?page=1&size=20&q=keywords:%22pectoral%20girdle%20morphometrics%20project%22; doi numbers are provided in Appendix S1: Table A1.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Abourachid, A. & Green, D.M. (1999) Origins of the frog-kick? Alternate-leg swimming in primitive frogs, families Leiopelmatidae and Ascaphidae. *Journal of Herpetology*, **33** (4), 657.

Adams, D.C. (2014) A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Systematic Biology*, **63** (5), 685–697.

Adams, D.C. (2016) Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. *Methods in Ecology and Evolution*, **7** (5), 565–572.

Adams, D.C. & Collyer, M.L. (2018) Phylogenetic ANOVA: group-clade aggregation, biological challenges, and a refined permutation procedure. *Evolution; international journal of organic evolution*, **72** (6), 1204–1215.

Adams, D.C., Collyer, M.L. & Kaliontzopoulou, A. (2020) *Geomorph: software for geometric morphometric analyses*. https://cran.r-project.org/package=geomorph.

Adler, D., Murdoch, D. & and others (2020) *Rgl: 3D visualization using OpenGL*. https://CRAN.R-project.org/package=rgl.

Ahn, A.N., Furrow, E. & Biewener, A.A. (2004) Walking and running in the red-legged running frog, *Kassina maculata*. *Journal of Experimental Biology*, **207** (Pt 3), 399–410.

Akella, T. & Gillis, G.B. (2011) Hopping isn't always about the legs: forelimb muscle activity patterns during toad locomotion. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **315** (1), 1–11.

Arnold, S.J. (1983) Morphology, performance and fitness. *American Zoologist*, **23** (2), 347–361.

Astley, H.C. (2016) The diversity and evolution of locomotor muscle properties in anurans. *The Journal of Experimental Biology*, **219** (19), 3163–3173.

Bigalke, R. (1927) Zur Myologie der Erdkröte. (*Bufo vulgaris*, Laurenti.). *Zeitschrift für Anatomie und Entwicklungsgeschichte*, **82** (1-3), 286–353.

Bijma, N.N., Gorb, S.N. & Kleinteich, T. (2016) Landing on branches in the frog *Trachycephalus resinifictrix* (Anura: Hylidae). *Journal of Comparative Physiology A*, **202** (4), 267–276.

Blaylock, L.A., Ruibal, R. & Platt-Aloia, K. (1976) Skin structure and wiping behavior of phyllomedusine frogs. *Copeia*, **1976** (2), 283.

Blotto, B.L., Pereyra, M.O., Faivovich, J., Dias, P.H.D.S. & Grant, T. (2017) Concentrated evolutionary novelties in the foot musculature of Odontophrynidae (Anura: Neobatrachia), with comments on adaptations for burrowing. *Zootaxa*, **4258** (5), 425–442.

Borcard, D., Gillet, F. & Legendre, P. (2011) *Numerical ecology with R*. Springer, New York.

Braus, H. (1919) *Der Brustschulterapparat der Froschlurche*. Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse: Abteilung B, Biologische Wissenschaften15. Winter, Heidelberg.

Brown, L.E. & Crespo, E.G. (2000) Burrowing behavior of the midwife toads *Alytes cisternasii* and *Alytes obstetricans* (Anura, Discoglossidae). *Alytes*, **17** (3-4), 101–113.

Brown, L.E., Jackson, H.O. & Brown, J.R. (1972) Burrowing behavior of the chorus frog, *Pseudacris streckeri*. *Herpetologica*, **28** (4), 325–328.

Burton, T.C. (1998) Are the distal extensor muscles of the fingers of anurans an adaptation to arboreality? *Journal of Herpetology*, **32** (4), 611.

Burton, T.C. (2001) Variation in the foot muscles of frogs of the family Myobatrachidae. *Australian Journal of Zoology*, **49** (5), 539.

Cardini, A. (2019) Integration and modularity in procrustes shape data: is there a risk of spurious results? *Evolutionary Biology*, **46** (1), 90–105.

Chadwell, B.A., Hartwell, H.J. & Peters, S.E. (2002) Comparison of isometric contractile properties in hindlimb extensor muscles of the frogs *Rana pipiens* and *Bufo marinus*: functional correlations with differences in hopping performance. *Journal of Morphology*, **251** (3), 309–322.

Choi, I. & Park, K. (1996) Variations in take-off velocity of anuran amphibians: relation to morphology, muscle contractile function and enzyme activity. *Comparative Biochemistry and Physiology Part A: Physiology*, **113** (4), 393–400.

Choi, I., Shim, J.H. & Ricklefs, R.E. (2003) Morphometric relationships of take-off speed in anuran amphibians. *Journal of experimental zoology. Part A, Comparative experimental biology*, **299** (2), 99–102.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F. & Ranzuglia, G. (2008) MeshLab: an open-source mesh processing tool. *Sixth Eurographics Italian Chapter conference: Salerno, Italy, July 2nd - 4th, 2008* (ed V. Scarano), pp. 129–136. Eurographics Assoc, Aire-la-Ville.

Citadini, J.M., Brandt, R., Williams, C.R. & Gomes, F.R. (2018) Evolution of morphology and locomotor performance in anurans: relationships with microhabitat diversification. *Journal of Evolutionary Biology*, **31** (3), 371–381.

Collyer, M.L. & Adams, D.C. (2018) RRPP: an R package for fitting linear models to high-dimensional data using residual randomization. *Methods in Ecology and Evolution*, **9** (7), 1772–1779.

Collyer, M.L. & Adams, D.C. (2020) *RRPP: linear model evaluation with randomized residuals in a permutation*. https://CRAN.R-project.org/package=RRPP.

Cox, S.M. & Gillis, G.B. (2015) Forelimb kinematics during hopping and landing in toads. *Journal of Experimental Biology*, **218** (19), 3051–3058.

Cundall, D., Fernandez, E. & Irish, F. (2017) The suction mechanism of the pipid frog, *Pipa pipa* (Linnaeus, 1758). *Journal of Morphology*, **278** (9), 1229–1240.

Currey, J.D. (1984) *The mechanical adaptations of bones*. Princeton University Press, Princeton, Guildford.

Davies, M. (1984) Osteology of the myobatrachine frog *Arenophryne rotunda* Tyler (Anura: Leptodactylidae) and comparisons with other myobatrachine genera. *Australian Journal of Zoology*, **32**, 789–802.

Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E. & Thelen, D.G. (2007) OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, **54** (11), 1940–1950.

Dryden, I.L. (2019) *shapes: statistical shape analysis*. https://CRAN.R-project.org/package=shapes.

Dutta, S.K. & Pradhan, B. (1985) Burrowing specialization in anurans. *Pranikee*, **6**, 41–52.

Emerson, S.B. (1976) Burrowing in frogs. *Journal of Morphology*, **149** (4), 437–458.

Emerson, S.B. (1978) Allometry and jumping in frogs: helping the twain to meet. *Evolution*, **32** (3), 551–564.

Emerson, S.B. (1979) The ilio-sacral articulation in frogs: form and function. *Biological Journal of the Linnean Scociety*, **11**, 153–168.

Emerson, S.B. (1983) Functional analysis of frog pectoral girdles. The epicoracoid cartilages. *Journal of Zoology*, **201** (3), 293–308.

Emerson, S.B. (1984) Morphological variation in frog pectoral girdles: testing alternatives to a traditional adaptive explanation. *Evolution*, **38** (2), 376–388.

Emerson, S.B. (1988) *Convergence and morphological constraint in frogs: variation in postcranial morphology: A contribution in celebration of the distinguished scholarship of Robert F. Inger on the occasion of his sixty-fifth birthday*. Fieldiana Zoology, **43**. Field Museum of Natural History, Chicago.

Emerson, S.B. (1990) Scaling of the epicoracoid horn muscle in arciferal frogs. *Journal of Herpetology*, **24** (1), 84–87.

Emerson, S.B. (1991) A biomechanical perspective on the use of forelimb length as a measure of sexual selection in frogs. *Journal of Evolutionary Biology*, **4** (4), 671–678.

Emerson, S.B. & Diehl, D. (1980) Toe pad morphology and mechanisms of sticking in frogs. *Biological Journal of the Linnean Society*, **13** (3), 199–216.

Emerson, S.B. & Koehl, M.A.R. (1990) The interaction of behavioral and morphological change in the evolution of a novel locomotor type: "flying" frogs. *Evolution*, **44** (8), 1931–1946.

Engelkes, K. (preprint, in review) Accuracy of bone segmentation and surface generation strategies analyzed by using synthetic CT volumes. http://www.doi.org/10.13140/RG.2.2.18389.86247.

Engelkes, K., Friedrich, F., Hammel, J.U. & Haas, A. (2018) A simple setup for episcopic microtomy and a digital image processing workflow to acquire high-quality volume data and 3D surface models of small vertebrates. *Zoomorphology*, **137** (1), 213–228.

Engelkes, K., Helfsgott, J., Hammel, J.U., Büsse, S., Kleinteich, T., Beerlink, A., Gorb, S.N. & Haas, A. (2019) Measurement error in µCT-based three-dimensional geometric morphometrics introduced by surface generation and landmark data acquisition. *Journal of Anatomy*, **235** (2), 357–378.

Enriquez-Urzelai, U., Montori, A., Llorente, G.A. & Kaliontzopoulou, A. (2015) Locomotor mode and the evolution of the hindlimb in Western Mediterranean anurans. *Evolutionary Biology*, **42** (2), 199–209.

Essner, R.L., Suffian, D.J., Bishop, P.J. & Reilly, S.M. (2010) Landing in basal frogs: evidence of saltational patterns in the evolution of anuran locomotion. *Die Naturwissenschaften*, **97** (10), 935–939.

Ferreira, R.C., Souza, A.A. de, Freitas, R.A., Campaner, M., Takata, C.S.A., Barrett, T.V., Shaw, J.J. & Teixeira, M.M.G. (2008) A phylogenetic lineage of closely related trypanosomes (Trypanosomatidae, Kinetoplastida) of anurans and sand flies (Psychodidae, Diptera) sharing the same ecotopes in brazilian amazonia. *The Journal of eukaryotic microbiology*, **55** (5), 427–435.

Frost, D.R. (2020) Amphibian Species of the World: an online reference. Version 6.1. http://research.amnh.org/herpetology/amphibia/index.html. Accessed 2 March 20.

Fürbringer, M. (1873) Zur vergleichenden Anatomie der Schultermuskeln. Teil 1. *Jenaische Zeitschrift für Medizin und Naturwissenschaft*, **7**, 237–320.

Gal, J.M. & Blake, R.W. (1987) Hydrodynamic drag of two frog species: *Hymenochirus boettgeri* and *Rana pipiens*. *Canadian Journal of Zoology*, **65** (5), 1085–1090.

Gaupp, E.W.T. (1896) *A. Ecker's und R. Wiedersheim's Anatomie des Frosches. Erste Abteilung. Lehre vom Skelett und vom Muskelsystem,* 3rd edn. Friedrich Vieweg und Sohn, Braunschweig.

Gillis, G.B. & Biewener, A.A. (2000) Hindlimb extensor muscle function during jumping and swimming in the toad (*Bufo marinus*). *Journal of Experimental Biology*, **203** (Pt 23), 3547–3563.

Gomes, F.R., Rezende, E.L., Grizante, M.B. & Navas, C.A. (2009) The evolution of jumping performance in anurans: morphological correlates and ecological implications. *Journal of Evolutionary Biology*, **22** (5), 1088–1097.

Gómez, R.O. & Lires, A.I. (2019) High ecomorphological diversity among Early Cretaceous frogs from a large subtropical wetland of Iberia. *Comptes Rendus Palevol*.

Gray, L.A., O'Reilly, J.C. & Nishikawa, K.C. (1997) Evolution of forelimb movement patterns for prey manipulation in anurans. *Journal of Experimental Zoology*, **277** (6), 417–424.

Griep, S., Schilling, N., Marshall, P., Amling, M., Hahne, L.M. & Haas, A. (2013) Pectoral girdle movements and the role of the glenohumeral joint during landing in the toad, *Rhinella marina* (Linnaeus, 1758). *Zoomorphology*, **132** (3), 325–338.

Gunz, P. & Mitteroecker, P. (2013) Semilandmarks: a method for quantifying curves and surfaces. *Hystrix, the Italian Journal of Mammalogy*, **24** (1), 103–109.

Haddad, C.F.B. & Sawaya, R.J. (2000) Reproductive modes of Atlantic forest hylid frogs: a general overview and the description of a new mode. *BIOTROPICA*, **32** (4), 862.

Herrel, A., Perrenoud, M., Decamps, T., Abdala, V., Manzano, A. & Pouydebat, E. (2013) The effect of substrate diameter and incline on locomotion in an arboreal frog. *Journal of Experimental Biology*, **216** (19), 3599–3605.

Herrel, A., van Damme, R., Vanhooydonck, B., Zaaf, A. & Aerts, P. (2000) Lizard locomotion: how morphology meets ecology. *Netherlands Journal of Zoology*, **50** (2), 261–277.

Hirano, M. & Rome, L.C. (1984) Jumping performance of frogs (*Rana pipiens*) as a function of muscle temperature. *Journal of Experimental Biology*, **108**, 429–439.

Hudson, N.J., Bennett, M.B. & Franklin, C.E. (2004) Effect of aestivation on long bone mechanical properties in the green-striped burrowing frog, *Cyclorana alboguttata*. *Journal of Experimental Biology*, **207** (3), 475–482.

Jones, E.I. (1933) Observations on the pectoral musculature of Amphibia Salientia. *Annals and Magazine of Natural History: Series 10*, **12** (70), 403–420.

Jorgensen, M.E. & Reilly, S.M. (2013) Phylogenetic patterns of skeletal morphometrics and pelvic traits in relation to locomotor mode in frogs. *Journal of Evolutionary Biology*, **26** (5), 929–943.

Kley, N.J. & Kearney, M. (2006) Adaptations for digging and burrowing. *Fins into Limbs* (ed B.K. Hall), pp. 284–309. University of Chicago Press.

Kumar, S., Stecher, G., Suleski, M. & Hedges, S.B. (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular biology and evolution*, **34** (7), 1812–1819.

Laurent, R.F. (1964) Adaptive modifications in frogs of an isolated highland fauna in Central Africa. *Evolution*, **18** (3), 458.

Lee, J. (2001) Evolution of a secondary sexual dimorphism in the toad, *Bufo marinus*. *Copeia*, **2001** (4), 928–935.

Liedvogel, M., Chapman, B.B., Muheim, R. & Åkesson, S. (2013) The behavioural ecology of animal movement: reflections upon potential synergies. *Animal Migration*, **1**, 39-46.

Maas, S.A., Ellis, B.J., Ateshian, G.A. & Weiss, J.A. (2012) FEBio: finite elements for biomechanics. *Journal of Biomechanical Engineering*, **134** (1), 11005.

Macarthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Sciences of the United States of America*, **43** (3), 293–295.

Manzano, A., Abdala, V. & Herrel, A. (2008) Morphology and function of the forelimb in arboreal frogs: specializations for grasping ability? *Journal of Anatomy*, **213** (3), 296–307.

Marcé-Nogué, J., Esteban-Trivigno, S. de, Escrig, C. & Gil, L. (2016) Accounting for differences in element size and homogeneity when comparing finite element models: Armadillos as a case study. *Palaeontologia Electronica*, **19.2.2T**, 1–22.

Menzies, J.I. & Tyler, M.J. (1977) The systematics and adaptations of some Papuan microhylid frogs which live underground. *Journal of Zoology*, **183** (4), 431–464.

Mielke, M. & Nyakatura, J.A. (2019) Bone microstructure in finite element modeling: the functional role of trabeculae in the femoral head of *Sciurus vulgaris*. *Zoomorphology*, **138** (4), 535–547.

Moen, D.S. (2019) What determines the distinct morphology of species with a particular ecology? The roles of many-to-one mapping and trade-offs in the evolution of frog ecomorphology and performance. *The American Naturalist*.

Moen, D.S., Irschick, D.J. & Wiens, J.J. (2013) Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. *Proceedings. Biological sciences / The Royal Society*, **280** (1773), 20132156.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (49), 19052–19059.

Nauwelaerts, S. & Aerts, P. (2006) Take-off and landing forces in jumping frogs. *The Journal of Experimental Biology*, **209** (1), 66–77.

Nauwelaerts, S., Ramsay, J. & Aerts, P. (2007) Morphological correlates of aquatic and terrestrial locomotion in a semi-aquatic frog, *Rana esculenta*: no evidence for a design conflict. *Journal of Anatomy*, **210** (3), 304–317.

Noble, G.K. & Jaeckle, M.E. (1928) The digital pads of the tree frogs. A study of the phylogenesis of an adaptive structure. *Journal of Morphology*, **45** (1), 259–292.

Nomura, F., Rossa-Feres, D.C. & Langeani, F. (2009) Burrowing behavior of *Dermatonotus muelleri* (Anura, Microhylidae) with reference to the origin of the burrowing behavior of Anura. *Journal of Ethology*, **27** (1), 195–201.

Oka, Y., Ohtani, R., Satou, M. & Ueda, K. (1984) Sexually dimorphic muscles in the forelimb of the Japanese toad, *Bufo japonicus*. *Journal of Morphology*, **180** (3), 297–308.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M. Henry H., Szoecs, E. & Wagner, H. (2019) *Vegan: community ecology package*. https://CRAN.R-project.org/package=vegan.

Oliver, J.A. (1951) "Gliding" in amphibians and reptiles, with a remark on an arboreal adaptation in the lizard, *Anolis carolinensis carolinensis* Voigt. *The American Naturalist*, **85** (822), 171–176.

Padhye, A.D., Jadhav, A., Sulakhe, S. & Dahanukar, N. (2015) Sexual dimorphism in the Kudremukh Bush Frog (Anura: Rhacophoridae: *Raorchestes tuberohumerus*) of the Western Ghats, India, with a note on its distribution and conservation status. *Journal of Threatened Taxa*, **7** (6), 7211–7222.

Paradis, E. & Schliep, K. (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.

Petrović, T.G., Vukov, T.D. & Kolarov, N.T. (2017) Sexual dimorphism in size and shape of traits related to locomotion in nine anuran species from Serbia and Montenegro. *Folia Zoologica*, **66** (1), 11–21.

Plate, T. & Heiberger, R. (2016) *Abind: combine multidimensional arrays*. https://CRAN.R-project.org/package=abind.

Pyron, A.R. & Wiens, J.J. (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, **61** (2), 543–583.

R Core Team (2019) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rand, A.S. (1952) Jumping ability of certain anurans, with notes on endurance. *Copeia*, **1952** (1), 15–20.

Reilly, S.M. & Jorgensen, M.E. (2011) The evolution of jumping in frogs: morphological evidence for the basal anuran locomotor condition and the radiation of locomotor systems in crown group anurans. *Journal of Morphology*, **272** (2), 149–168.

Reilly, S.M., Montuelle, S.J., Schmidt, A., Krause, C., Naylor, E. & Essner, R.L. (2016) Functional evolution of jumping in frogs: interspecific differences in take-off and landing. *Journal of Morphology*, **277** (3), 379–393.

Reynaga, C.M., Astley, H.C. & Azizi, E. (2018) Morphological and kinematic specializations of walking frogs. *Journal of experimental zoology. Part A, Ecological and integrative physiology*, **329** (2), 87–98.

Ricklefs, R.E. & Miles, D.B. (1994) Ecological and evolutionary inferences from morphology: an ecological perspective. *Ecological morphology: Integrative organismal biology* (eds P.C. Wainwright & S.M. Reilly), pp. 13–41. The University of Chicago Press, Chicago, London.

Ritland, R.M. (1955) Studies on the post-cranial morphology of *Ascaphus truei*. II. Myology. *Journal of Morphology*, **97** (2), 215–282.

Robovská-Havelkova, P., Aerts, P., Roček, Z., Prikryl, T., Fabre, A.-C. & Herrel, A. (2014) Do all frogs swim alike? The effect of ecological specialization on swimming kinematics in frogs. *The Journal of Experimental Biology*, **217** (Pt 20), 3637–3644.

RStudio Team (2018) *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA.

Sanders, J. & Davies, M. (1983) Burrowing behaviour and associated hindlimb myology in some Australian hylid and leptodactylid frogs. *The Australian Zoologist*, **21** (2), 123–142.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P. & Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, **9** (7), 676–682.

Schlager, S. (2017) Morpho and Rvcg - shape analysis in R. *Statistical shape and deformation analysis: Methods, implementation and applications* (eds Z. Guoyan, L. Shuo & S. Gabor), pp. 217–256. Academic Press, Amsterdam.

Schlosser, G. (2002) Modularity and the units of evolution. *Theory in biosciences = Theorie in den Biowissenschaften*, **121** (1), 1–80.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, **9** (7), 671–675.

Sherman, M.A., Seth, A. & Delp, S.L. (2013) What is a moment arm? Calculating muscle effectivness in biomechanical models using generalized coordinates. *Proceedings of the ASME Design Engineering Technical Conferences*, **7B**, V07BT10A052.

Sherratt, E., Vidal-García, M., Anstis, M. & Keogh, J.S. (2017) Adult frogs and tadpoles have different macroevolutionary patterns across the Australian continent. *Nature ecology & evolution*, **1** (9), 1385–1391.

Soliz, M., Tulli, M.J. & Abdala, V. (2017) Osteological postcranial traits in hylid anurans indicate a morphological continuum between swimming and jumping locomotor modes. *Journal of Morphology*.

Trueb, L. (1973) Bones, Frogs, and evolution. *Evolutionary biology of the anurans: Contemporary research on major problems* (ed J.L. Vial), pp. 65–132. University of Missouri Press, Columbia.

Trueb, L. & Gans, C. (1983) Feeding specializations of the Mexican burrowing toad, *Rhinophrynus dorsalis* (Anura: Rhinophrynidae). *Journal of Zoology*, **199** (2), 189–208.

Vera, M.C., Ferretti, J.L., Abdala, V. & Cointry, G.R. (2020) Biomechanical properties of anuran long bones: correlations with locomotor modes and habitat use. *Journal of Anatomy*.

Vidal-García, M., Byrne, P.G., Roberts, J.D. & Keogh, J.S. (2014) The role of phylogeny and ecology in shaping morphology in 21 genera and 127 species of Australo-Papuan myobatrachid frogs. *Journal of Evolutionary Biology*, **27** (1), 181–192.

Videler, J.J. & Jorna, J.T. (1985) Functions of the sliding pelvis in *Xenopus laevis*. *Copeia*, **1985** (1), 251.

Wainwright, P.C. (1994) Functional morphology as a tool in ecological research. *Ecological morphology: Integrative organismal biology* (eds P.C. Wainwright & S.M. Reilly), pp. 42–59. The University of Chicago Press, Chicago, London.

Wainwright, P.C. (2007) Functional versus morphological diversity in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, **38** (1), 381–401.

Wainwright, P.C., Alfaro, M.E., Bolnick, D.I. & Hulsey, C.D. (2005) Many-to-one mapping of form to function: a general principle in organismal design? *Integrative and Comparative Biology*, **45** (2), 256–262.

Wang, Z., Ji, A., Endlein, T., Samuel, D., Yao, N., Wang, Z. & Dai, Z. (2014) The role of fore- and hindlimbs during jumping in the Dybowski's frog (*Rana dybowskii*). *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **321** (6), 324–333.

Wells, K.D. (2007) *The ecology & behavior of amphibians*. University of Chicago Press, Chicago.

Wilson, M.P., Espinoza, N.R., Shah, S.R. & Blob, R.W. (2009) Mechanical properties of the hindlimb bones of bullfrogs and cane toads in bending and torsion. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology*, **292** (7), 935–944.

Zug, G.R. (1971) Buoyancy, locomotion, morphology of the pelvic girdle and hindlimb, and systematics of cryptodiran turtles. *Miscellaneous Publications Museum of Zoology University of Michigan*, **142** (1-98).

Zug, G.R. (1972) Anuran locomotion: structure and function. I. preliminary observations on relation between jumping and osteometrics of appendicular and postaxial skeleton. *Copeia*, **1972** (4), 613–624.

Zug, G.R. (1978) Anuran locomotion-structure and function 2: jumping performance of semiaquatic, terrestrial and arboreal frogs. *Smithsonian Contributions to Zoology*, **276**, 1–31.

Zug, G.R. (1985) Anuran locomotion: fatigue and jumping performance. *Herpetologica*, **41** (2), 188–194.

Appendix S1

[The following information can be found in Appendix S1 at the end of this thesis, pp. 255-265.]

FIGURE A1 Plot of principal components 3 and 4 of overall species mean shapes of pectoral girdle bones (landmark dataset i).

TABLE A1 Specimens, locomotor modes, and CT scanning parameter.

TABLE A2 Specimens, locomotor modes, and MorphoSource media number.

TABLE A3 Definition of landmarks and curves of semilandmarks (fixed landmarks adopted from Engelkes et al., 2019).

- Chapter five -

A simple setup for episcopic microtomy and a digital image processing workflow to acquire high-quality volume data and 3D surface models of small vertebrates

Karolin Engelkes[1], Frank Friedrich[2], Jörg U. Hammel[3,4], Alexander Haas[1]

1) Center of Natural History (CeNak), Universität Hamburg, Germany
2) Biocenter Grindel, Universität Hamburg, Germany
3) Institute of Materials Research, Helmholtz-Zentrum Geesthacht, Germany
4) Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller-Universität Jena, Germany

Published in Zoomorphology (2018),

volume 137, issue 1, pages 213-228, doi 10.1007/s00435-017-0386-3.

Contributions of KE:

Study design: ~ 90%

Data collection: 100%

Data analysis: 100%

Manuscript drafting: 100%

Contributions confirmed by supervisor: _____

Prof. Dr. Alexander Haas

Abstract

The use of volume data and digital three-dimensional (3D) surface models in biology has increased quickly and steadily. Various methods are available to acquire 3D data, among them episcopic imaging techniques. Based on the episcopic microscopy with on-block staining protocol of Weninger et al. (Anat Embryol 197:341–348, 1998), we describe a simple and versatile setup for episcopic microtomy. It is composed of a consumer DSLR digital camera combined with standard histology equipment. The workflow of block surface staining and imaging, image processing, stack alignment, surface generation (including a custom Amira® macro), and 3D model editing is described in detail. For our sample specimen (*Alytes obstetricans*; Amphibia: Anura) we obtained images with a pixel size of 5.67 x 5.67 µm^2. The generated image stacks allowed distinguishing different tissues and were well-suited for creating a 3D surface model. We analyzed the alignment quality achieved by various selections of specimen and fiducial marker spots. The fiducial spots had a significant positive effect on the alignment quality with the best alignment having a maximum mean alignment error of about 44.7 µm. We further tested the APS-C camera with combinations of macro lens, extension tube or teleconverter. The macro lens and extension tube yielded the smallest pixel size of 2.53 x 2.53 µm^2. Considering data quality and resolution, and depending on object sizes and research goals, DSLR captured episcopic microtomy can be an alternative to other techniques, such as traditional histological sectioning or micro-computed tomography.

Keywords: Block surface imaging, On-block staining, Alignment, Volume data, Surface model

Introduction

The use of digital three-dimensional (3D) volume representations and surface models for analysis and visualization increased substantially within the past two decades (e.g., Joschko et al. 1991; Haas 2001; Golding et al. 2009; Staedler et al. 2013; Lautenschlager et al. 2014; Gan et al. 2016; Heiss et al. 2016; Henne et al. 2017). 3D surface models have also been used increasingly in geometric morphometrics (e.g., Gunz et al. 2012; Piras et al. 2015; Werneburg et al. 2015), X-ray reconstruction of moving morphology (e.g., Brainerd et al. 2010; Griep et al. 2013), multibody dynamic analyses and musculoskeletal modeling (e.g., Kargo and Rome 2002; Curtis et al. 2010; Charles et al. 2016), finite element analyses (e.g., Cox et al. 2011; Kleinteich et al. 2012; Fortuny et al. 2015), and computational fluid dynamics (e.g., Hammel et al. 2013; Rahman et al. 2015; also see Davies et al. 2017 and references therein).

In vertebrate morphology, arguably one of the first procedures to generate 3D models of biological structures was introduced by Born (1876, 1883). He used stacked wax plates, cut out according to histological sections projected on them via a prism, to build a physical, enlarged 3D representation of amphibian nasal structures. Later progress in computer technology allowed creating digital 3D models of biologically relevant structures by surface

scanning, by photogrammetry, or based on image stacks and volume data. Methods of volume data acquisition include micro-computed tomography (μCT), magnetic resonance imaging (MRI), confocal laser-scanning microscopy (CLSM), or digitally stacked serial images of a specimen. A review of these and related techniques and the pertinent literature has been presented by Geyer et al. (2009). For the stacked image approach, digitized histological serial sections (e.g., Handschuh et al. 2010) and episcopic 3D imaging methods (i.e., methods that generate volume data by repetitively digitizing the cut surface of a histologically embedded specimen during its sectioning; also see Geyer et al. 2009) have been used.

Weninger et al. (1998) described an episcopic 3D imaging method to acquire high-resolution volume data. They mounted a block with an embedded specimen on a microtome for sectioning and combined *in situ* staining of the cut surface (Hegre and Brashear 1946, 1947), with computer-assisted block surface digitizing and image processing techniques (Laan et al. 1989; Odgaard et al. 1990). Unlike traditional histological sections on slides, the main advantages of the episcopic imaging are that the resulting image stacks are gapless, rapidly generated, not affected by geometric distortions, and inherently aligned such that the natural outline of the specimen is preserved. Weninger et al. (1998) produced two stacks of gray-scale images, one stack of unstained and one of stained block surfaces images, respectively. Due to the precise congruence of the cross section images, automatic image subtraction between stained and unstained images was possible and removed shining-through background structures (Weninger et al. 1998; Fig. 1). They used the resulting stack of subtracted images for the generation of a surface model. In their setup with highly precise image capture position, post-sectioning alignment was unnecessary. If choosing digitized histological sections on slides (without positional precision in image capture) to create volume data, however, image alignment is an obligatory, yet an error-prone and time-consuming step in the preparation of volume data sets. Such alignment can result in unnatural shapes; for example, wrongly straightened curved structures (the so-called “banana” problem), false *z*-axis orientation, or reconstruction of a symmetric shape from an unsymmetrical specimen (Streicher et al. 1997; Malandain et al. 2004).

The originally published setup for episcopic microscopy with on-block staining has proven valuable for the acquisition of high-quality volume data (Weninger et al. 1998). It, however, seemed rather complex: a dissection microscope with a ring-light was placed perpendicular to the cut surface of the block, and a video-camera connected to a computer was operated by a custom software macro for image acquisition. Later efforts pushed episcopic microscopy to higher resolution (high-resolution episcopic microscopy, HREM, and

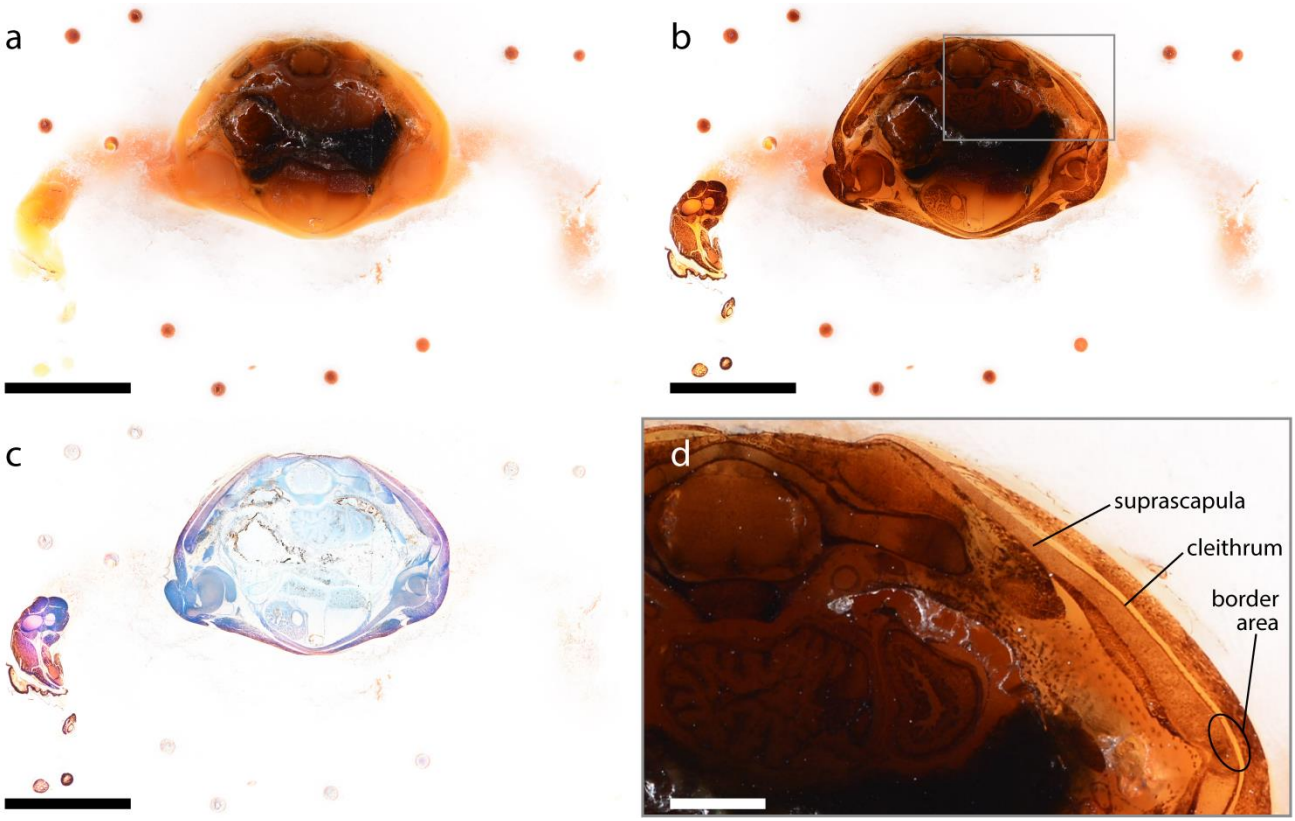


Fig. 1 Cross section through the pectoral girdle of *Alytes obstetricans* (ZMH A12442). Episcopic images (cropped) of unstained (**a**) and stained (**b**) cut surface of the paraffin block. **c** Result of inverted difference of images **a** and **b**. **d** Detail of **b** showing no clear border of cleithrum. Black scales: 5 mm; white scale: 1 mm

related episcopic imaging methods; Weninger et al. 2006; Mohun and Weninger 2012a, b; Geyer et al. 2014). HREM is a fluorescence-based, automated approach to episcopic microscopy that requires a specialized setup and plastic resin embedding; both limiting its application to certain laboratory settings and object sizes. Automation in sectioning and image capturing is certainly highly desirable for high-throughput applications but also requires higher investment in specialized technical equipment not needed in many other applications.

Our approach herein aims to use standard equipment commonly present in histology laboratories, to assemble a setup that is simple, straightforward, and versatile without sacrificing resolution. In the following, we propose a setup for episcopic microtomy with block surface staining that harnesses the high resolution of contemporary DSLR cameras. We test the performance of the setup and describe a workflow of image acquisition and processing to create high-quality surface models. Our approach neither strives for full automation nor for perfectly aligned image stacks; rather, time investment in manual operation of the microtome and image acquisition, and digital alignment replaced more complex hardware solutions. We applied fiducial markers to counter alignment artifacts and statistically assessed the alignment quality.

Materials and methods

Specimen and embedding

A museum specimen of *Alytes obstetricans* (Laurenti, 1768) (Amphibia: Anura; ZMH A12442; snout-vent-length 28 mm) was chosen as test specimen. The head, arms, and trunk of the specimen were partly skinned to allow for a better light transmission deep into the sample. The further steps of paraffin (Paraplast Plus®; Leica Biosystems) embedding were adopted from Weninger et al. (1998). Most importantly the specimen was impregnated with lead ions prior to paraffin infiltration. The lead ions accumulate in the tissues and allow for later *in situ* staining of the cut surface of the block (Hegre and Brashear 1946, 1947). The protocol is given in Online Resource 1.

Setup for episcopic microtomy

The paraffin block was mounted (melted) onto a wooden base block and clamped into a manually driven rotatory microtome (Microm™ HM 340 E; Microm International GmbH). The block was oriented to ensure transverse sectioning of the specimen. The fastened block was moved to a position close to the upper turning point of the microtome. This zero position was marked on the microtome hand wheel for subsequent image acquisition. We used a position slightly before peak block position in order to have space for adjustments of the area captured by the camera in case the area of interest shifted relative to the camera view during sectioning.

A 24 megapixel digital APS-C sensor camera (Nikon D7200®; Nikon Corporation) equipped with a macro lens (Nikon AF-S VR Micro-Nikkor® 105 mm 1:2,8G IF-ED) was mounted on a focusing rack (Manfrotto® 454), which in turn was fastened to a tripod (Gitzo® G2220). The optical axis of the camera was leveled perpendicular to the block surface. The camera alignment was done by eye, as this proved to be precise enough in previous episcopic imaging trials (unpubl. data). A UV filter and lens hood were mounted to protect the lens from chemicals and paraffin shavings, to enhance contrast, and to prevent lens flare.

One Nikon SB-R200 Wireless Speedlight was placed on each side of the microtome at level with the specimen block in its zero position. Flashes were triggered in TTL mode by the built-in flash of the camera as commander. Flash position was slightly behind the cut surface to mostly illuminate deep into the block. Careful placement of the flashes was essential to reduce glare and reflections from the cut block surface and to lighten up dark areas behind the

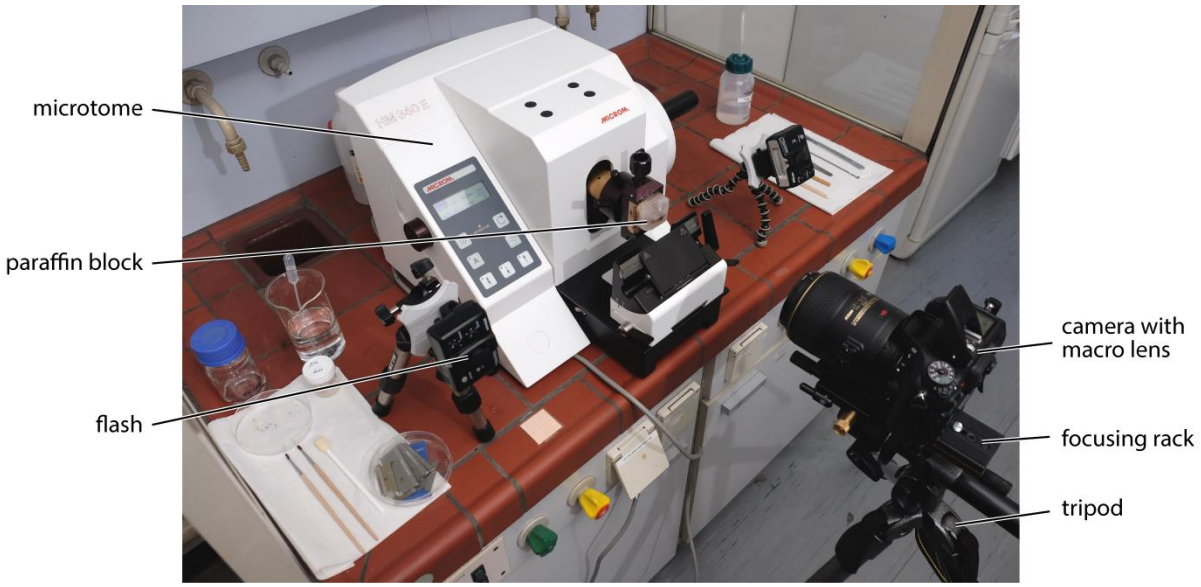


Fig. 2 Setup for episcopic microtomy, lens hood removed

cut surface plane. This backlit effect enhanced the clarity of the region of interest in the block surface plane. The setup is shown in Fig. 2.

We operated the camera in manual focus mode with constant magnification in order to achieve identical pixel sizes (i.e., real world pixel equivalents) for all images. The lens magnification was adjusted such that the field of view covered an area big enough for the expected maximum cross section area of the specimen to fit in, without adding too much empty space around it. The aim was to maximize areal coverage of the specimen projection on the camera sensor and, thus, maximize specimen spatial resolution. Metric-grid paper placed in the focus plane helped to estimate the proper field of view adjustment.

Block sectioning and imaging

Focusing on the cut block surface was done manually with the focusing rack (lens settings/magnification kept constant). A self-timer and a mirror pre-release were set for the camera to reduce vibrations during image capture. Once all camera settings were set, an image of grid-paper placed in the focus plane was taken (uncompressed RAW format [NEF]) to allow for pixel size calibration.

Fiducial markers were applied to the cut block surface: multiple holes were punched into the block with a pin to create markers for later image alignment (also see, e.g., Malandain et al. 2004 and references therein). Holes were placed in each corner of the field of view using a wood block with a tight, perpendicular drilled hole as guide for the inserted pin. To counter slight deviations from perpendicularity the holes in the upper right and lower left corner were applied with the same guide block orientation, whereas those in the other corners were given

the opposite orientation (woodblock turned 180° around the long axis of the pin). The holes were filled with melted dark chocolate (non-toxic, excellent cutting properties) to gain high contrast to the paraffin block. This resulted in subcircular chocolate spots around the specimen on the images. At the beginning, four holes were made into the paraffin, one in each corner. The holes were up to 18 mm deep and new holes were added as needed. As it was not possible to fill the entire length of holes with chocolate, but just about 1 mm at once, chocolate had to be added during sectioning. Additional holes were applied for redundancy, in case the chocolate content in another hole ended before it was refilled; in total up to six holes per corner were created.

During block sectioning, images of each pristine cut surface were taken first. Then, a solution of sodium sulfide (Na_2S; 7%) was applied to the cut surface for about 20 s using a sponge applicator. The sodium sulfide reacted with the lead ions accumulated in tissues and stained tissues of the specimen dark against the surrounding unchanged paraffin (Hegre and Brashear 1946, 1947). The degree of staining was controlled visually. When the surface was considered to be dark enough, superfluous solution was removed with another sponge applicator. Then the stained cut surface was photographed in unchanged *x*–*y* position so that the images of the unstained and stained condition were perfectly congruent. The break of the microtome could have been used to fix the block in its position during staining and image acquisition, but this was not necessary as the sponge applicator imposed only minor forces on the cut surface so that the block did not move between unstained and stained images. After taking this pair of images, three sections (10 µm thickness each) of the block surface were cut off. The block was then carefully set to zero position and the next image pair was captured. Care was taken to keep pictures of unstained and stained block surface always alternating in the sequence of images, because this facilitates subsequent automated sorting of images into separate stacks. The specimen was only sectioned in the region of interest, here the anterior part of the body. In this particular project, we saved sections on glass slides and stained them with Azan stain (Mulisch and Welsch 2010, with modifications) for future reference. In the pectoral girdle region, the section before the sodium-sulfide-stained and imaged section was saved; anterior to the pectoral girdle, sections were saved at lager intervals as this region was not of particular interest. If it was not possible to save the intended section, e.g., because the section got damaged during mounting, the section before or after (a sodium-sulfide-stained one) it was saved. Two stained histological sections were digitized for comparison with episcopic images (Leica DM6000 B; Leica Microsystems GmbH). The images were processed (cropping, background cleaning, adjusting of brightness and contrast) in Fiji (based

on ImageJ 1.51j; Schindelin et al. 2012; Schneider et al. 2012) and arranged in Adobe® Illustrator® CS6 (version 16.0.3; Adobe® Systems Software).

Digital image processing and pre-alignment

Image RAW files were converted to TIFF in IrfanView® (version 4.41; Irfan Skiljan, http://www.irfanview.com). The images of unstained and stained block surfaces were opened as separate stacks in Fiji. The size of the pixels was calculated based on the image of metric grid-paper. The differences of corresponding images (stained/unstained pairs) from the two stacks were calculated with Fiji's *Image Calculator* and the resulting images were inverted. This inverted stack shall be called 'subtracted stack' from here on. Brightness and contrast of the image stacks were adjusted in Fiji.

The stack of stained images was opened in Amira® (version 6.0.1; Konrad-Zuse-Zentrum Berlin, FEI Visualization Sciences Group) and aligned using the automatic alignment function of the *Align Slices* module (*least squares alignment* algorithm). Subsequently, the automatic alignment was checked and corrected manually where needed. The alignment transformations were saved and applied to align the subtracted stack as well. By doing so, the subtracted stack was aligned in exactly the same way as the stained stack and corresponding images in both stacks were kept in identical *x*–*y* positions. This first alignment was applied to allow the use of the *Magic Wand* tool with the *All slices* option for the segmentation of the fiducial spots in later processing steps. These 'pre-aligned' stained and subtracted stacks were then converted to 8-bit gray-scale in Fiji to be compatible with the *Segmentation Editor* of Amira®.

Segmentation, alignment, and assessment of alignment quality

We selected one fiducial marker in each corner of the aligned stained gray-scale stack (markers I–IV). These selected four markers comprised the spots that later served to test the alignment quality. For each corner, the one fiducial marker was chosen that had the best quality spots (most circular and sharply bordered) and a maximum number of spots through the stack. The spots from the chosen markers were segmented from the aligned gray-scale stained stack as separate *materials* in Amira® (*Magic Wand* tool with constantly adjusted thresholds and *Brush* tool). Marker spots that had no sharp border or were incomplete were omitted during segmentation. This resulted in 164–371 segmented spots per marker canal. The segmentations were stored in a *LabelField* (subsequently called 'test-spot *LabelField*') in Amira®. The marker's respective spot groups will be referred to as spot groups i–iv (Fig. 3).

The marker spots of the remaining markers and the specimen were segmented as separate *materials* (*Magic Wand* tool with *All slices* option and selection growing) into a new *LabelField*. Using this *LabelField* and the pre-aligned gray-scale stained stack, we created five image stacks with different compositions of elements in their original gray values (*Arithmetic* module; Fig. 3): (1) the specimen only, (2) the specimen and all marker spots except those resulting from markers I–IV, (3) all marker spots except those resulting from markers I–IV, (4) all marker spots of the upper left and lower right corner except those resulting from markers I–IV (marker spots from makers I–IV were included where needed to bridge those images, which lacked other appropriate spots), and (5) one marker spot per corner (two spots in one corner, when we had to switch between markers) not resulting from markers I–IV (marker spots from makers I–IV were included where needed to bridge those images, which lacked other appropriate spots).

These five stacks containing different structure compositions were each aligned in Amira® (alignments 1–5; automatic *least squares alignment* algorithm; manually corrected if needed). The transformations of each alignment were then applied to a copy of the test-spot *LabelField* that contained the segmentations of spot groups i–iv. The resulting five differently aligned test-spot *LabelFields* were exported as tif-image stacks and the midpoint coordinates (x, y, and z) of each spot in groups i–iv were determined using a custom Fiji script resulting in 20 clouds of midpoints (5 alignments with 4 spot groups each).

We used the first principal component of each cloud of midpoints as line of best fit to approximate the long axis of the respective fiducial markers. For each observed point of a given midpoint cloud, the shortest distance of it to the fitted line was calculated. The resulting sets of distances will be called 1.i, 1.ii, through 5.iv depending on the alignment and spot group they belong to. The distance sets were tested for normality (Shapiro–Wilk test). This test did not support the assumption of normality for all sets but two (2.v, 3.iii). A Kruskal–Wallis rank sum tests was performed to test for significant differences between the distance sets. A Dunn's tests with a p-value adjustment according to Holm was run to reveal pairwise differences between distance sets, because the Kruskal–Wallis test indicated significant differences. Line fittings, distance calculations, and statistical tests were performed in R (version 3.3.2; R Core Team 2016) using various functions of the *stats* package, and the Dunn's test implemented in the *PMCMR* package (Pohlert 2014) via RStudio® (version 1.0.136; RStudio Team 2016). A p-value smaller than 0.05 was considered significant for all tests The transformations from the alignment that resulted in the statistically best result (alignment 5) were then used to re-align the pre-aligned gray-scale stacks of the stained and

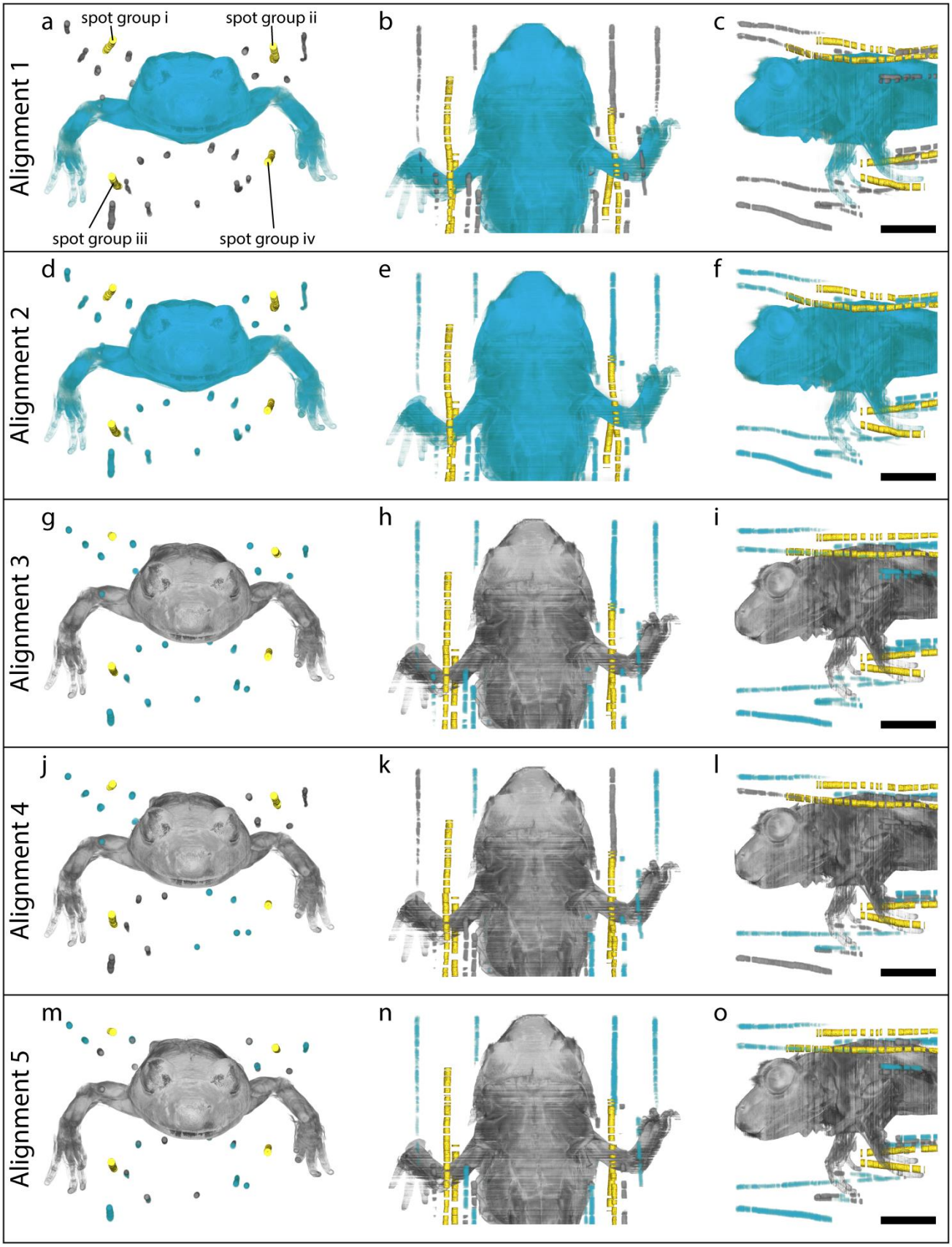


Fig. 3 3D visualizations of differently aligned stained image stacks (converted to gray-scale) with the structures belonging to different alignments highlighted. Views from left to right: anterior, dorsal, lateral. Scale: 5 mm; all views at same scale. Blue: volume rendering of structure used for alignment; yellow: *Isosurface* of test-spot segmentations aligned accordingly; gray: volume rendering of remaining structures. **a–c**: Alignment 1. **d–f**: Alignment 2. **g–i**: Alignment 3. **j–l**: Alignment 4. **m–o**: Alignment 5

subtracted stacks in Amira®. The steps of data alignment and statistical analyses are illustrated in Fig. 4. Skeletal elements of the right pectoral girdle were segmented (*Brush* tool with interpolation of up to 5 images) in the re-aligned, final gray-scale stacks. We used both, the stained and the subtracted stack, for segmentation, switching back and forth between them. A scale (5 mm of the grid-paper image) and selected additional skeletal elements were segmented without distinguishing between bone and cartilage.

Surface generation and processing

The segmented structures were exported out of Amira® as separate polygon surfaces (obj-files) using a custom Amira® macro (see *MultiExport* macro and macro documentation in Online Resource 2): all voxel belonging to a given structure/*material* were extracted (*Arithmetic* module) and a *Isosurface* was created. Subsequently, a polygon surface was computed (*ExtractSurface* module). A copy of the surface was exported (subsequently called 'original surface') and it was reduced in polygon count (*SimplificationEditor*) and smoothed (*SmoothSurface* module). Polygon count reduction and smoothing were done in small iterative steps to prevent artifacts in the surface. Finally, the simplified surface was exported.

Subsequent steps of surface processing were adopted from established workflows (e.g., Kleinteich et al. 2008; Friedrich et al. 2015; Mekonen et al. 2015; Gan et al. 2016). The simplified surfaces were imported into MeshLab (version 1.3.3; Visual Computing Lab–ISTI–CNR; http://meshlab.sourceforge.net/) and further reduced in polygon count (*Quadric Edge Collapse Decimision*), smoothed (*Taubin Smooth*), and cleaned (*Compact faces*, *Compact vertices*, *Merge Close Vertices*, *Remove Duplicate Faces*, *Remove Duplicated Vertex*). Surfaces were then passed to MODO® (version 10.1v2; The Foundry; see, e.g., Ablan 2008). The shape of each simplified surface was manually corrected (*Move*, *Smooth*, *Surface Pen*) using the corresponding original surface as template, as smoothing and polygon count reduction altered the shape of the simplified surface and caused artifacts such as holes. The original surface of the scale bar was replaced by an accordingly scaled cube. The cube was positioned in a way that it had the same distance to the camera as the center of the 3D model and that one face was orthogonal to the virtual camera for rendering visualizations. The final surfaces were assigned *materials* (colors) and the scene was illuminated with *Directional* and *Dome Lights*. The rendered images were arranged and labeled in Adobe® Illustrator® CS6 including the replacement of the scale cube by a bar of equivalent length.

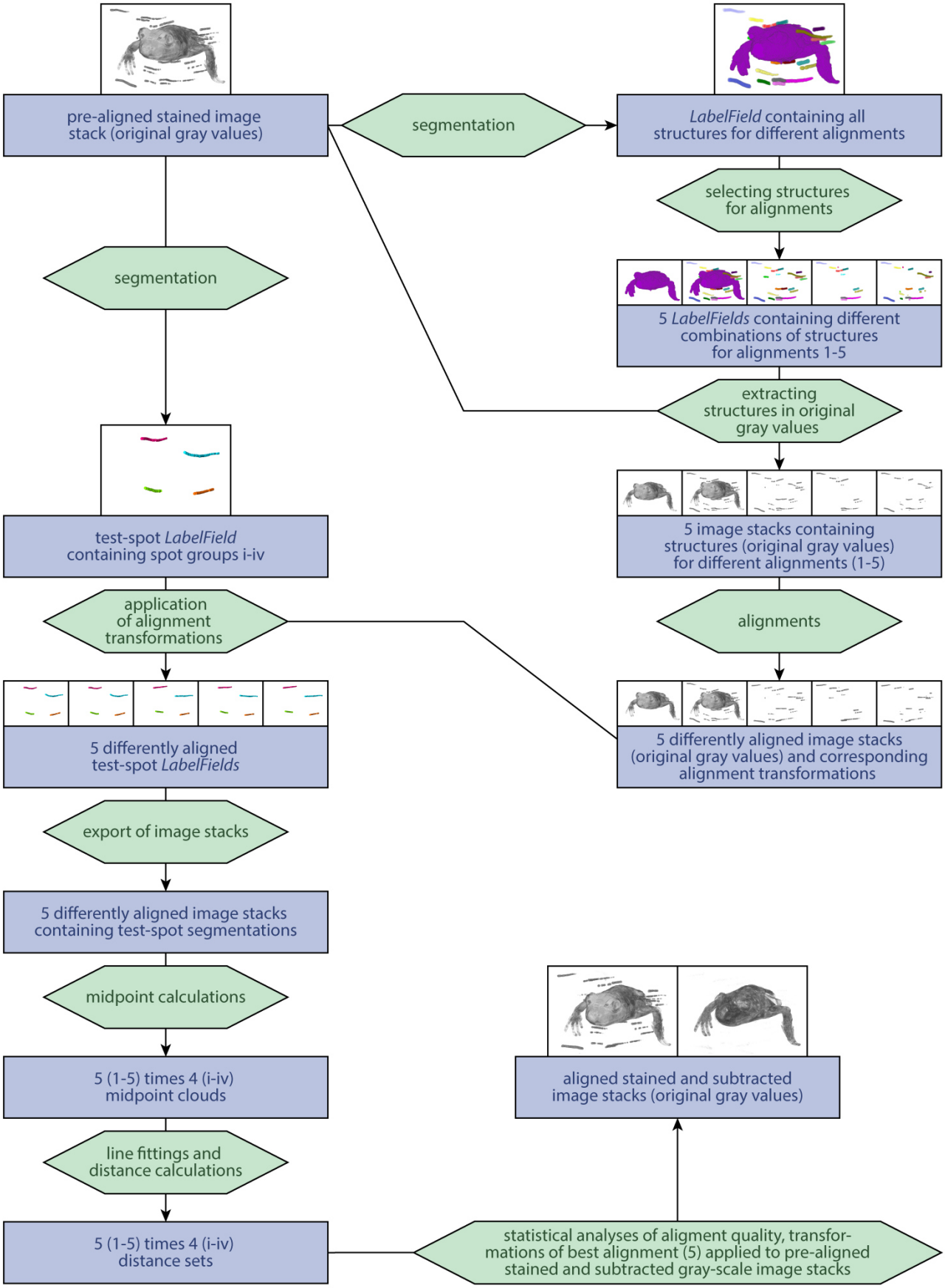


Fig. 4 Steps of data alignment and statistical analysis of alignment quality

Pixel sizes

We tested different settings of the macro lens and distances of the camera's sensor plane to metric grid-paper to calculate pixel sizes and field of views captured by the camera. We further tested the effect of using a teleconverter (AF-S, TC-14E III; Nikon) and an extension tube (36 mm; Kenko Tokina Co., Ltd.) with the lens because these pieces of equipment are commonly used by photographers and readily available. The pixel size was determined eight times for each distance and lens configuration and the mean pixel size was calculated. Measurements and calculations were done by a custom Fiji macro that required the user to mark a known distance on each grid-paper image eight times.

Results

Episcopic image stacks and alignment quality

The partial sectioning of the *Alytes* specimen resulted in stacks of 620 episcopic images plus one scale image (grid-paper). Image acquisition took about 45 h including the time for mounting the sections on glass slides. The images had a pixel size of 5.67 µm resulting in a voxel size of 5.67 µm in *x* and *y* and 30 µm in *z*. There was some variation in staining intensity among images and within the same image. This resulted in banding when the images were rendered as a volume (Fig. 3), but had no negative effect on the segmentation or the surfaces creation process (Fig. 5). Structures in dark areas of the stained stack (central body parts of the specimen) often showed better contrast in the subtracted stack (Fig. 1). The obtained image quality allowed us to well distinguish cartilage and bone tissues, except for the suprascapula cartilage that is partly covered by the very thin cleithrum bone (Fig. 1d); the edges of the cleithrum were not traceable with certainty. It, however, was still possible to roughly segment both structures by comparing subsequent images and approximating borders (Fig. 5a, b). In addition to the skeletal structures, muscle tissue, tendons, major nerves, and other organs were clearly distinguishable.

Using different compositions of structures for the stack alignment resulted in different alignment qualities, with statistically significant differences between most distance sets. There, however, was no significant difference between corresponding distance sets of both alignments in which the specimen was part of the structures used for the alignment (1 and 2). There further were no significant differences between distance sets 3.i and 5.i, 3.ii and 5.ii, 3.iii and 4.iii, 3.iv and 4.iv, 3.iv and 5.iv, and 4.iv and 5.iv, respectively, and some other pairs. Box-and-whisker plots of the different distance sets are given in Fig. 6; descriptive statistics

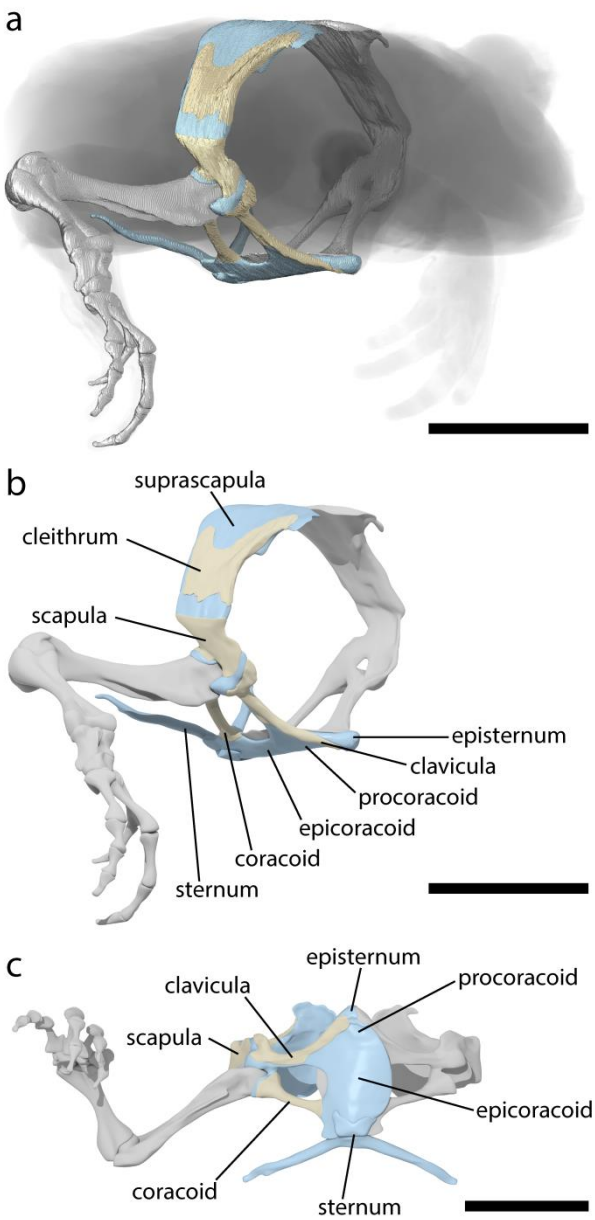


Fig. 5 a Unprocessed *Isosurfaces* of the segmented structures (pectoral girdle and additional structures) in a volume rendering (converted to uniform gray-scale) of the *Alytes obstetricans* specimen (ZMH A12442; image created in Amira®). The surfaces appear stepped due to an inter-image distance of 30 µm. Anterolateral view, anterior to the right. Beige: bone; blue: cartilage; gray: additional skeleton without distinguishing bone and cartilage. Scale: 5 mm. **b**, **c** Simplified, smoothed, and corrected surface model of the surfaces in a (image rendered in MODO®). Anterolateral (**b**) and ventral (**c**) views. Same color code as in **a**, scales 5 mm

of the distance sets and the results of Kruskal–Wallis and Dunn's tests are provided in Online Resources 3 and 4, respectively.

Histological serial sections

The digitized histological sections had a pixel size of 1.67 µm; smaller pixel sizes would have been achievable with higher magnifications at the Leica slide scanner. The lead impregnation had no visible effect on the tissue integrity and the quality of the Azan staining (Fig. 7a). In cases where a sodium-sulfide-stained section was saved the colors of Azan staining appeared slightly more brownish than those of non-sodium-sulfide-stained sections (not quantified; Fig. 7b, d). The histological sections showed all anatomical details usually visible in Azan stained sections. For some of the sections geometric distortions were obvious or some parts of the specimen got damaged or lost (Fig. 7a, b).

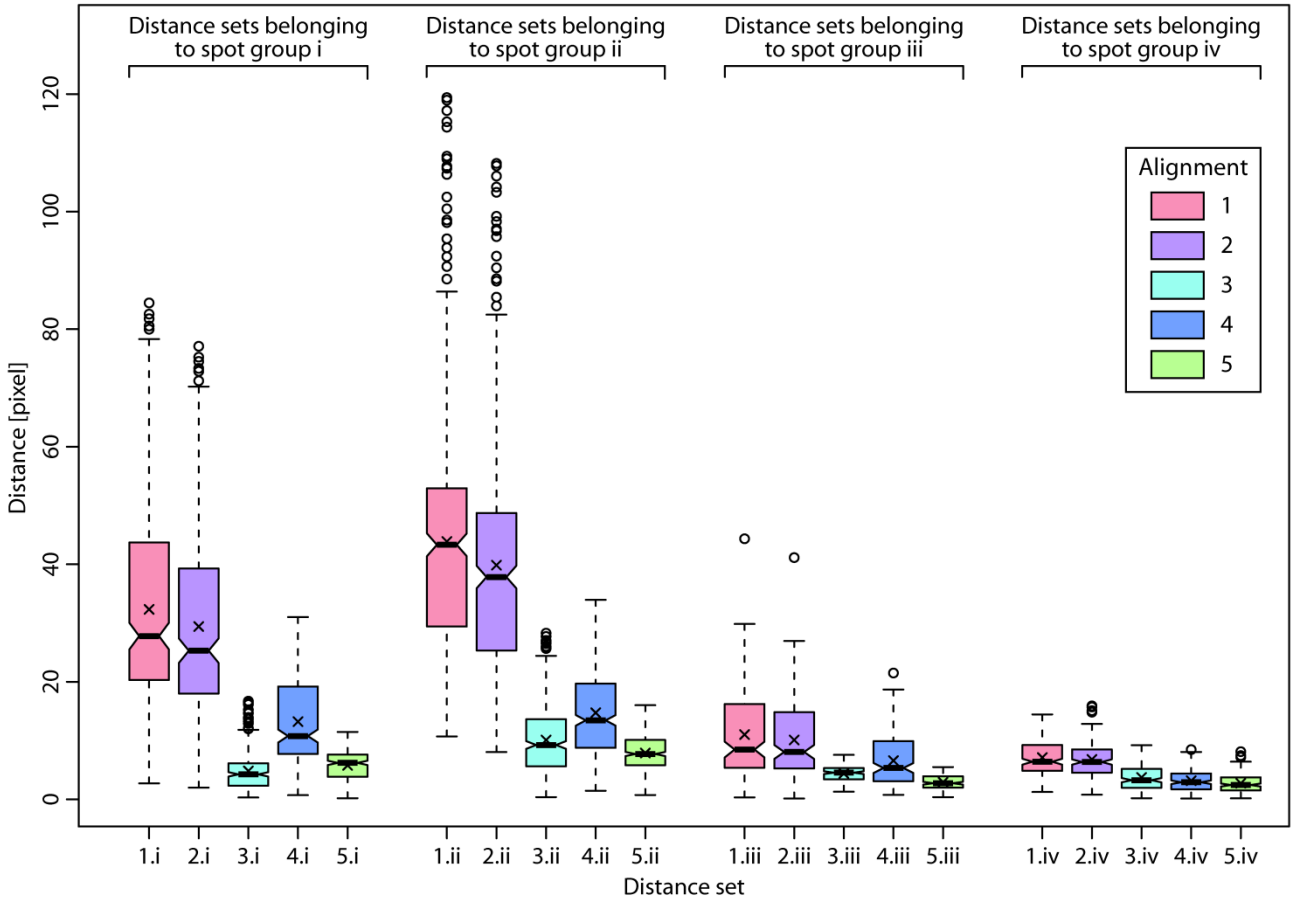


Fig. 6 Notched box-and-whisker plots of distance sets grouped by the respective spot group they belong to. If notches of two boxes do not overlap, this is strong evidence that there is a true difference in the means of the respective distance sets (McGill et al. 1978). Circle: potential outlier; cross: mean

Pixel sizes

The smallest calculated pixel size (2.53 µm) was obtained using the macro lens with the extension tube. The associated area captured by the camera was 15.19 mm by 10.13 mm. Other setup variations with their respective pixel sizes and fields of view are given in Table 1.

Discussion

Limitations, improvements, and strengths

Our intention was to devise a lab setup for episcopic microscopy with photographic and histological equipment readily accessible in many labs. For our test case, a medium sized frog, the setup worked very well. However, size of the object matters. The minimum and maximum size of a specimen intended for episcopic microtomy is limited by the specifications of the equipment used for sectioning and digital image capture. In our protocol, the specimen is embedded in paraffin and the block has to fit the microtome used. Previously,

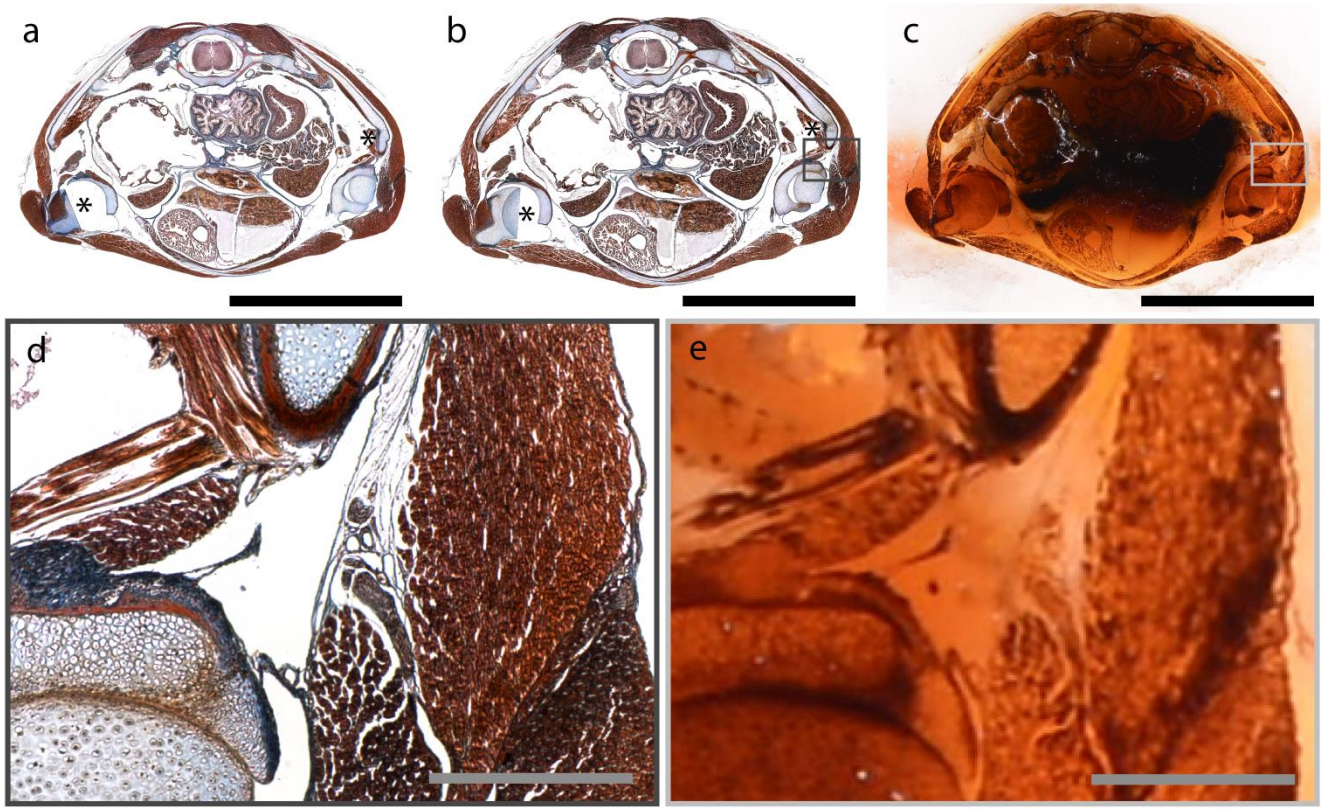


Fig. 7 Histological sections (**a**, **b**, **d**) and episcopic images (**c**, **e**) of *Alytes obstetricans* (ZMH A12442). **a** Digitized Azan stained histological section that was not contrasted with sodium sulfide. **b** Digitized Azan stained histological section that was previously contrasted with sodium sulfide. Sections of **a** and **b** were located on the same glass slide and thus Azan-stained identically. **c** Sodium sulfide contrasted episcopic image that shows the same section as in **b**. **d** Detail of **b** showing different tissues and some brownish remains of the sodium sulfide staining. **e** Detail of **c** showing the resolution limit of the episcopic images. Asterisk: regions damaged during sectioning, section mounting, or Azan-staining; black scales: 5 mm; gray scales: 0.5 mm

we sectioned a much bigger toad with a similar protocol, however, on a sliding microtome and the camera above the microtome (unpubl. data). For very large objects sectioning or milling can be a demanding task (e.g., Visible Human Project®; Spitzer and Whitlock 1998). Also, the increasing field of view in large objects scales inversely with the spatial resolution. Camera sensor resolution, lens magnification, and lens resolution are to be considered for small specimens. Small objects can be accommodated by inserting extension tubes, teleconverters, or macro bellows (not tested herein) between camera body and lens to increase the magnification; another option is a lens with already high magnification (e.g., Canon MP-E 65mm f/2.8 1-5x Macro Photo). However, at some point the crystalline fine structure of the paraffin will impose a limit at the lower end of the range; according to our experience we recommend different protocols (resin embedding) for vertebrate objects that are substantially smaller in diameter and for which small pixel sizes are needed (0.25–0.5 µm x 0.25–0.5 µm; Weninger et al. 2006).

Table 1 Influence of focus plane to sensor distance, extension tube, and teleconverter on pixel size (first value) and field of view (second).

focus plane to sensor distance [mm]	Macro lens only	Macro lens and extension tube	Macro lens and teleconverter
310	3.68 µm* 22.07 x 14.71 mm^2	-	-
320	4.19 µm 25.16 x 16.78 mm^2	-	-
324	-	2.53 µm* 15.19 x 10.13 mm^2	-
330	4.70 µm 28.21 x 18.80 mm^2	2.97 µm 17.80 x 11.87 mm^2	-
334	-	-	2.63 µm* 15.77 x 10.51 mm^2
340	5.19 µm 31.14 x 20.76 mm^2	3.80 µm 22.79 x 15.20 mm^2	2.85 µm 17.12 x 11.41 mm^2
350	5.65 µm 33.91 x 22.60 mm^2	4.23 µm 25.36 x 16.90 mm^2	3.19 µm 19.13 x 12.76 mm^2
360	6.10 µm 36.57 x 24.38 mm^2	4.78 µm 28.70 x 19.14 mm^2	3.54 µm 21.25 x 14.16 mm^2

Asterisk denotes smallest pixel size (highest magnification) obtainable with respective equipment. Values were rounded to the nearest two decimals

Increasing the sensor's pixel density (decreasing the sensor's pixel pitch, i.e., the distance between sensor photo sites) increases the potential camera resolution given the lens has enough resolving power. Note that APS-C DSLR sensors commonly have a higher pixel density than most full frame (FX) sensors and, therefore, an APS-C sensor camera was the better choice for our setup.

Our microtome did not allow the arresting of the block in precisely the same position for consecutive images. Therefore, we obtained stacks of unaligned images. Perfectly aligned stacks can be achieved with other microtomes. We, however, purposely wanted to test our setup and post-sectioning protocols with a microtome type that is more likely to be encountered in many laboratories. Furthermore, we continued using a tripod (easily accessible), although care must be taken that the tripod does not move relative to the specimen block during sectioning. The camera could have been coupled mechanically to the microtome by some device, but we did not want to give up simplicity. Rather, we demonstrated that episcopic imaging can be well done with a tripod alone.

Block surface staining with sodium sulfide does not produce consistent staining intensities (Geyer et al. 2009), which, in turn, lead to an inhomogeneous volume (Fig. 3). With regard to structure segmentation, this turned out to be advantageous, because some structures were better discernible in lightly stained and others in strongly stained images.

Despite the limitations of the setup, we obtained image stacks that, in our opinion, are of a high quality, and very well suitable for morphological work on specimens of this size class. The advantages of episcopically captured image stacks are obvious (Weninger et al. 1998;

Geyer et al. 2009): The digitalized cross sections are undistorted, no sections are lost due to processing problems, and the corresponding images in the stained and unstained stacks are congruent allowing for image subtraction and use of both stacks for segmentation in 3D processing software.

Alignment quality

Our setup requires the virtual alignment of stacked images. Fiducial markers have been shown to be valuable for image stack registration and to overcome known alignment problems such as the "banana" problem (e.g., Toga and Arnicar 1985; Brändle 1989; Ford-Holevinski et al. 1991; Humm et al. 1995; Goldszal et al. 1996; Streicher et al. 1997; Rau et al. 2013). Our results confirmed that the use of external markers can increase alignment quality.

We found no significant difference in the alignment quality of the two alignments (alignments 1–2) that used, among others, the specimen itself for the alignment. When comparing these two alignments, the alignment including the marker spots (2) generally results in better, although not significantly better values (i.e., the distances of the marker spot midpoints to the fitted lines were on average smaller). This clearly demonstrated the influence of the marker spots on the alignment quality. The influence, however, was small possibly because the image area occupied by the marker spots was small compared to the area of the specimen, and both together served for alignment in alignment 2. In other words, the specimen had a higher weight during the alignment process. All alignments that exclusively used various subsets of the marker spots (alignments 3–5) performed statistically better than the two including the specimen (1–2). This indicates that the marker spots provided important alignment information that helped to reduce alignment artifacts.

The alignment using one fiducial marker spot per corner (5) resulted in the statistically best alignment quality; the alignment using all marker spots (3) resulted in second best. The difference between the qualities of these two alignments is based on the significant difference of distance sets 3.iii and 5.iii only. This indicates that both these alignments do not differ much in their quality.

In alignment 5, the mean alignment error (in terms of mean distance of midpoints to, respectively, fitted lines) of spot groups iii–iv was smaller than 3 pixel; this equals about 17 µm and was smaller than the image stack resolution in *z*. The highest mean alignment error of this alignment was in spot group ii and equalled 44.7 µm, approximately 1.5 times the resolution in *z*. The data indicate that there is a slight deviation between the natural shape of the specimen and its' shape after automated alignment. Whether this alignment error is

relevant or not depends on the kind of subsequent analysis and needs to be determined. Episcopic microtomy datasets of our setup may not be suitable for analyses that require high-precision measurements such as (geometric) morphometrics.

We observed significant differences between most distance sets within each alignment. This indicates that different parts of the images are aligned with different accuracies. Possible explanations for this observation could be as follows: the marker spots considered for the alignment had different areas and thereby different weights during the alignment process; the holes driven into the block deviated non-uniformly from perpendicularity and led to a shift of the *z*-axis.

We did not analyze the accuracy with which the coordinates of the midpoints were determined. Some errors could have been introduced during the segmentation of the spots chosen to test the alignment quality and during the determination of their midpoint coordinates. Yet, we expect the possible error in midpoint determination to have only minor effect on the alignment comparisons as we used the same test-spot *LabelField* for all distance calculations.

Alternative steps and software

There is a large, if not bewildering, number of software packages that can be freely combined to achieve the results desired. Herein we present the solution that suited us best, but alternative packages and workflows do exist. Amira® offers various functions of image processing that we performed in Fiji instead. For surface generation, there is an alternative Amira® module (*Generate Surface*); a custom script for exporting multiple separate surfaces created with *Generate Surface* is available from the authors. The Fiji extension TrakEM2 (Cardona et al. 2012) can perform alignment and segmentation tasks. VGStudio MAX® (Volume Graphics GmbH) also offers some functions similar to Amira®. Separate image stacks could have been exported for each segmented structure (image stack export included in the macro in Online Resource 2). The resulting image stacks could have been used to create surfaces in, for example, Imaris® (Bitplain) or GOM Inspect (GOM GmbH). GOM Inspect is also capable of polygon count reduction, smoothing, and cleaning of the surfaces. For final surface processing and image rendering Maya® (Autodesk) or Blender® (Blender Foundation) are alternative options.

Comparison to other methods of generating volume data

For acquiring volume data, specimens that are suitable for episcopic microtomy commonly will be suitable for other imaging techniques as well. One approach is histological serial sectioning with subsequent mounting on slides, staining, digitizing, stacking, and alignment (e.g., Handschuh et al. 2010). Microscopic slides can be examined at very high magnifications and contrast between tissues of interests can be enhanced with appropriate staining methods; the episcopic microtomy workflow suggested herein has certain limitations in tissue staining and resolution (also compare Fig. 7). The major disadvantage of histological sections with regard to digital 3D reconstruction is that the sections are more or less distorted during the sectioning and mounting process. Episcopic microtomy does not have this problem because the block surface, not the cut section, is digitized. It might be interesting to evaluate the applicability of non-rigid alignment algorithms to elastically register a digitized histological section to a corresponding episcopic image, for example, by adapting the method suggested by Saalfeld et al. (2012) (also see Laan et al. 1989).

Micro-computed tomography is a widely used imaging method in biology (Neues and Epple 2008; Mizutani and Suzuki 2012). It allows the non-destructive acquisition of isotropic, high-resolution (sub-micrometer) volume data. µCT data are inherently aligned. For soft tissue visualization (e.g., muscle), contrast staining may be required (e.g., Metscher 2009a, b; Gignac et al. 2016). Using a contrast agent, however, is no longer non-destructive because the agent might change tissue properties and possibly remains in the specimens for some time (Schmidbaur et al. 2015 and references therein). Even with contrast staining, making cartilage discernible and distinguishable from other tissues remains difficult (unpubl. data). Histological serial sections and episcopic microtomy are suitable methods for soft tissue and cartilage visualization and may be a better choice than µCT scans depending on both the research question and considerations about the integrity of the specimens.

Magnetic resonance imaging has been used as another non-destructive technique to acquire inherently aligned volume data of specimens (e.g., Arnold et al. 2000; Turnbull and Mori 2007; Driehuys et al. 2008; Gabbay-Benziv et al. 2017). MRI data acquisition is typically performed without physically altering or staining the specimen (but see Rohrer et al. 2005). The resolution, however, is still limited and dependent on magnetic field strength. To our knowledge, voxel sizes down to an edge length of 10 µm have been achieved (Lee et al. 2015). High-resolution MRI scanners are much less accessible to most researchers than µCT scanners or even histological equipment.

Micro-computed tomography and MRI are volume data acquisition methods that have several advantages with regard to shape and specimen preservation, and with regard to speed of data acquisition (several hours versus several days to weeks in episcopic microtomy and histological serial sectioning). µCT and MRI, however, currently have limits in tissue contrast (one color channel) and visibility. Thus it depends on the research goal, which method is most efficient with regard to time and the data quality needed.

Some studies combined different imaging methods (e.g., Laan et al. 1989; Pieles et al. 2007; Schulz et al. 2012; Handschuh et al. 2013; Herdina et al. 2015). This adds advantages and potentially neutralizes the disadvantages of each technique. In our labs, we combined µCT scans (for bones) with subsequent episcopic microtomy (for cartilage and soft tissues) of the same specimen; data sets could be registered to each other (unpubl. data).

Conclusion

Episcopic microtomy workflows can produce huge amounts of high-quality digital images with excellent differential tissue contrast, at a small inter-image distance, and without the need for physical storage space for glass slides. Selected sections of a specimen can still be preserved on glass slides during the episcopic process if necessary. After applying standard staining (e.g., Azan), these histological sections provide additional anatomical information. Our setup is clearly not suitable for HREM. If resolutions at the cellular level are needed or objects are much smaller than ours, stacking of histological sections (Handschuh et al. 2013) or HREM (Weninger et al. 2006) need to be considered. Furthermore, our setup is not recommended for projects that need high automation or high throughput. The described setup and procedures, however, can easily be set up by the occasional user to take advantage of episcopic imaging. Our simple episcopic microtomy setup allowed relatively fast specimen processing and the resulting datasets offered ample resolution. Episcopic microtomy has a lot of potential in morphological work pipelines to generate volume data.

Acknowledgements

We thank Anna-Maria Vogt and Lena Schwinger for skillful help with the lab-work. We also thank the students who have worked with episcopic microtomy data in our class or during their bachelor projects and helped gaining experiences with the workflow, in particular Stine Griep, Lisa Malina Hahne, Vivienne-Christine Ahrens, Juliana Stahl, and Mehria Sedik. We also thank the anonymous reviewers for their helpful comments on the manuscript.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or living animals performed by any of the authors.

References

Ablan D (2008) The official luxology modo guide. Version 301. Course Technology, Cengage Learning, Boston

Arnold AS, Salinas S, Asakawa DJ, Delp SL (2000) Accuracy of muscle moment arms estimated from MRI-based musculoskeletal models of the lower extremity. Comput Aided Surg 5:108–119. https://doi.org/10.1002/1097-0150(2000)

Born G (1876) Ueber die Nasenhöhlen und den Thränennasengang der Amphibien. Morphol Jahrb 2:577–646

Born G (1883) Die Plattenmodelliermethode. Arch f mikrosk Anat 22:584–599

Brainerd EL, Baier DB, Gatesy SM, Hedrick TL, Metzger KA, Gilbert SL, Crisco JJ (2010) X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. J Exp Zool A Ecol Genet Physiol 313A:262–279. https://doi.org/10.1002/jez.589

Brändle K (1989) A new method for aligning histological serial sections for three-dimensional reconstruction. Comput Biomed Res 22:52–62. https://doi.org/10.1016/0010-4809(89)90015-3

Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, Tomancak P, Hartenstein V, Douglas RJ, Samuel A (2012) TrakEM2 software for neural circuit reconstruction. PLoS ONE 7:e38011. https://doi.org/10.1371/journal.pone.0038011

Charles JP, Cappellari O, Spence AJ, Wells DJ, Hutchinson JR (2016) Muscle moment arms and sensitivity analysis of a mouse hindlimb musculoskeletal model. J Anat 229:514–535. https://doi.org/10.1111/joa.12461

Cox PG, Fagan MJ, Rayfield EJ, Jeffery N (2011) Finite element modelling of squirrel, guinea pig and rat skulls: using geometric morphometrics to assess sensitivity. J Anat 219:696–709. https://doi.org/10.1111/j.1469-7580.2011.01436.x

Curtis N, Jones MEH, Evans SE, Shi J, O'Higgins P, Fagan MJ (2010) Predicting muscle activation patterns from motion and anatomy: modelling the skull of *Sphenodon* (Diapsida: Rhynchocephalia). J R Soc Interface 7:153–160. https://doi.org/10.1098/rsif.2009.0139

Davies TG, Rahman IA, Lautenschlager S, Cunningham JA, Asher RJ, Barrett PM, Bates KT, Bengtson S, Benson RBJ, Boyer DM, Braga J, Bright JA, Claessens LPAM, Cox PG, Dong X, Evans AR, Falkingham PL, Friedman M, Garwood RJ, Goswami A, Hutchinson JR, Jeffery NS, Johanson Z, Lebrun R, Martínez-Pérez C, Marugán-Lobón J, O'Higgins PM, Metscher B, Orliac M, Rowe TB, Rücklin M, Sánchez-Villagra MR, Shubin NH, Smith SY, Starck JM, Stringer C, Summers AP, Sutton MD, Walsh SA, Weisbecker V, Witmer LM, Wroe S, Yin Z, Rayfield EJ, Donoghue PCJ (2017) Open

data and digital morphology. Proc. R. Soc. B 284:20170194. https://doi.org/10.1098/rspb.2017.0194

Driehuys B, Nouls J, Badea A, Bucholz E, Ghaghada K, Petiet A, Hedlund LW (2008) Small animal imaging with magnetic resonance microscopy. ILAR J 49:35–53. https://doi.org/10.1093/ilar.49.1.35

Ford-Holevinski TS, Castle MR, Herman JP, Watson SJ (1991) Microcomputer-based three-dimensional reconstruction of in situ hybridization autoradiographs. J Chem Neuoanat 4:373–385. https://doi.org/10.1016/0891-0618(91)90044-D

Fortuny J, Marcé-Nogué J, Heiss E, Sanchez M, Gil L, Galobart À (2015) 3D bite modeling and feeding mechanics of the largest living amphibian, the Chinese Giant salamander *Andrias davidianus* (Amphibia: Urodela). PLoS ONE 10:e0121885. https://doi.org/10.1371/journal.pone.0121885

Friedrich F, Schulz J, Kubiak M, Beckmann F, Wilde F (2015) The larval head anatomy of *Rhyacophila* (Rhyacophilidae) with discussion on mouthpart homology and the groundplan of Trichoptera. J Morphol 276:1505–1524. https://doi.org/10.1002/jmor.20475

Gabbay-Benziv R, Reece EA, Wang F, Bar-Shir A, Harman C, Turan OM, Yang P, Turan S (2017) A step-wise approach for analysis of the mouse embryonic heart using 17.6Tesla MRI. Magn Reson Imaging 35:46–53. https://doi.org/10.1016/j.mri.2016.08.008

Gan LL, Hertwig ST, Das I, Haas A (2016) The anatomy and structural connectivity of the abdominal sucker in the tadpoles of *Huia cavitympanum*, with comparisons to *Meristogenys jerboa* (Lissamphibia: Anura: Ranidae). J Zool Syst Evol Res 54:46–59. https://doi.org/10.1111/jzs.12113

Geyer SH, Mohun TJ, Weninger WJ (2009) Visualizing vertebrate embryos with episcopic 3D imaging techniques. ScientificWorldJOURNAL 9:1423–1437. https://doi.org/10.1100/tsw.2009.154

Geyer SH, Nöhammer MM, Mathä M, Reissig L, Tinhofer IE, Weninger WJ (2014) High-Resolution Episcopic Microscopy (HREM): a tool for visualizing skin biopsies. Microsc Microanal 20:1356–1364. https://doi.org/10.1017/S1431927614013063

Gignac PM, Kley NJ, Clarke JA, Colbert MW, Morhardt AC, Cerio D, Cost IN, Cox PG, Daza JD, Early CM, Echols MS, Henkelman RM, Herdina AN, Holliday CM, Li Z, Mahlow K, Merchant S, Müller J, Orsbon CP, Paluh DJ, Thies ML, Tsai HP, Witmer LM (2016) Diffusible iodine-based contrast-enhanced computed tomography (diceCT): an emerging tool for rapid, high-resolution, 3-D imaging of metazoan soft tissues. J Anat 228:889–909. https://doi.org/10.1111/joa.12449

Golding RE, Ponder WF, Byrne M (2009) Three-dimensional reconstruction of the odontophoral cartilages of Caenogastropoda (Mollusca: Gastropoda) using micro-CT: Morphology and phylogenetic significance. J Morphol 270:558–587. https://doi.org/10.1002/jmor.10699

Goldszal AF, Tretiak OJ, Liu DD, Hand PJ (1996) Multimodality multidimensional image analysis of cortical and subcortical plasticity in the rat brain. Ann Biomed Eng 24:430–439. https://doi.org/10.1007/BF02660892

Griep S, Schilling N, Marshall P, Amling M, Hahne LM, Haas A (2013) Pectoral girdle movements and the role of the glenohumeral joint during landing in the toad, *Rhinella marina* (Linnaeus, 1758). Zoomorphology 132:325–338. https://doi.org/10.1007/s00435-013-0189-0

Gunz P, Ramsier M, Kuhrig M, Hublin J, Spoor F (2012) The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. J Anat 220:529–543. https://doi.org/10.1111/j.1469-7580.2012.01493.x

Haas A (2001) Mandibular arch musculature of anuran tadpoles, with comments on homologies of amphibian jaw muscles. J Morphol 247:1–33. https://doi.org/10.1002/1097-4687(200101)247

Hammel JU, Jahn H, Möckel L, Herzen J, Greving I, Wilde F, Beckmann F, Nickel M (2013) Computational fluid dynamics of sponge aquiferous systems. HASYLAB Annual Report

Handschuh S, Schwaha T, Metscher BD (2010) Showing their true colors: a practical approach to volume rendering from serial sections. BMC Dev Biol 10:41. https://doi.org/10.1186/1471-213X-10-41

Handschuh S, Baeumler N, Schwaha T, Ruthensteiner B (2013) A correlative approach for combining microCT, light and transmission electron microscopy in a single 3D scenario. Front Zool 10:44. https://doi.org/10.1186/1742-9994-10-44

Hegre ES, Brashear AD (1946) Block-surface staining. Stain Technol 21:161–164. https://doi.org/10.3109/10520294609110063

Hegre ES, Brashear AD (1947) The block-surface method of staining as applied to the study of embryology. Anat Rec 97:21–28. https://doi.org/10.1002/ar.1090970103

Heiss E, Handschuh S, Aerts P, van Wassenbergh S (2016) Musculoskeletal architecture of the prey capture apparatus in salamandrid newts with multiphasic lifestyle: does anatomy change during the seasonal habitat switches? J Anat:757–770. https://doi.org/10.1111/joa.12445

Henne S, Friedrich F, Hammel JU, Sombke A, Schmidt-Rhaesa A (2017) Reconstructing the anterior part of the nervous system of *Gordius aquaticus* (Nematomorpha, Cycloneuralia) by a multimethodological approach. J Morphol 278:106–118. https://doi.org/10.1002/jmor.20623

Herdina AN, Plenk H, Benda P, Lina PHC, Herzig-Straschil B, Hilgers H, Metscher BD (2015) Correlative 3D-imaging of *Pipistrellus* penis micromorphology: validating quantitative microCT images with undecalcified serial ground section histomorphology. J Morphol 276:695–706. https://doi.org/10.1002/jmor.20372

Humm JL, Macklis RM, Lu XQ, Yang Y, Bump K, Beresford B, Chin LM (1995) The spatial accuracy of cellular dose estimates obtained from 3D reconstructed serial tissue autoradiographs. Phys Med Biol 40:163–180

Joschko M, Graff O, Müller PC, Kotzke K, Lindner P, Pretschner DP, Larink O (1991) A non-destructive method for the morphological assessment of earthworm burrow systems in three dimensions by X-ray computed tomography. Biol Fertil Soils 11:88–92. https://doi.org/10.1007/BF00336369

Kargo WJ, Rome LC (2002) Functional morphology of proximal hindlimb muscles in the frog *Rana pipiens*. J Exp Biol 205:1987–2004

Kleinteich T, Beckmann F, Herzen J, Summers AP, Haas A (2008) Applying x-ray tomography in the field of vertebrate biology: form, function, and evolution of the skull of caecilians (Lissamphibia: Gymnophiona). Proc SPIE 7078:70780D. https://doi.org/10.1117/12.795063

Kleinteich T, Maddin HC, Herzen J, Beckmann F, Summers AP (2012) Is solid always best? Cranial performance in solid and fenestrated caecilian skulls. J Exp Biol 215:833–844. https://doi.org/10.1242/jeb.065979

Laan AC, Lamers WH, Huijsmans DP, Kortschot AT, Smith J, Strackee J, Los JA (1989) Deformation-corrected computer-aided three-dimensional reconstruction of immunohistochemically stained organs: application to the rat heart during early organogenesis. Anat Rec 224:443–457. https://doi.org/10.1002/ar.1092240313

Lautenschlager S, Bright JA, Rayfield EJ (2014) Digital dissection – using contrast-enhanced computed tomography scanning to elucidate hard- and soft-tissue anatomy in the Common Buzzard *Buteo buteo*. J Anat 224:412–431. https://doi.org/10.1111/joa.12153

Lee CH, Blackband SJ, Fernandez-Funez P (2015) Visualization of synaptic domains in the *Drosophila* brain by magnetic resonance microscopy at 10 micron isotropic resolution. Sci Rep 5:1234168. https://doi.org/10.1038/srep08920

Malandain G, Bardinet É, Nelissen K, Vanduffel W (2004) Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations. NeuroImage 23:111–127. https://doi.org/10.1016/j.neuroimage.2004.04.038

McGill R, Tukey JW, Larsen WA (1978) Variations of box plots. Am Stat 32:12. https://doi.org/10.2307/2683468

Mekonen HK, Hikspoors JPJM, Mommen G, Köhler SE, Lamers WH (2015) Development of the ventral body wall in the human embryo. J Anat 227:673–685. https://doi.org/10.1111/joa.12380

Metscher BD (2009a) MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. BMC Physiol 9:1–14. https://doi.org/10.1186/1472-6793-9-11

Metscher BD (2009b) MicroCT for developmental biology: a versatile tool for high-contrast 3D imaging at histological resolutions. Dev Dyn 238:632–640. https://doi.org/10.1002/dvdy.21857

Mizutani R, Suzuki Y (2012) X-ray microtomography in biology. Micron 43:104–115. https://doi.org/10.1016/j.micron.2011.10.002

Mohun TJ, Weninger WJ (2012a) Episcopic three-dimensional imaging of embryos. Cold Spring Harb Protoc 2012:641–646. https://doi.org/10.1101/pdb.top069567

Mohun TJ, Weninger WJ (2012b) Generation of volume data by episcopic three-dimensional imaging of embryos. Cold Spring Harb Protoc 2012:681–682. https://doi.org/10.1101/pdb.prot069591

Mulisch M, Welsch U (eds) (2010) Romeis Mikroskopische Technik, 18th edn. Spektrum Akademischer Verlag, Heidelberg

Neues F, Epple M (2008) X-ray microcomputer tomography for the study of biomineralized endo- and exoskeletons of animals. Chem Rev 108:4734–4741. https://doi.org/10.1021/cr078250m

Odgaard A, Andersen K, Melsen F, Gundersen HJ (1990) A direct method for fast three-dimensional serial reconstruction. J Microsc 159:335–342. https://doi.org/10.1111/j.1365-2818.1990.tb03038.x

Pieles G, Geyer SH, Szumska D, Schneider J, Neubauer S, Clarke K, Dorfmeister K, Franklyn A, Brown SD, Bhattacharya S, Weninger WJ (2007) μMRI-HREM pipeline for high-throughput, high-resolution phenotyping of murine embryos. J Anat 211:132–137. https://doi.org/10.1111/j.1469-7580.2007.00746.x

Piras P, Sansalone G, Teresi L, Moscato M, Profico A, Eng R, Cox TC, Loy A, Colangelo P, Kotsakis T (2015) Digging adaptation in insectivorous subterranean eutherians. The enigma of *Mesoscalops montanensis* unveiled by geometric morphometrics and finite element analysis. J Morphol 276:1157–1171. https://doi.org/10.1002/jmor.20405

Pohlert T (2014) The pairwise multiple comparison of mean ranks package (PMCMR). R package. http://CRAN.R-project.org/package=PMCMR

Rahman IA, Darroch SAF, Racicot RA, Laflamme M (2015) Suspension feeding in the enigmatic Ediacaran organism *Tribrachidium* demonstrates complexity of Neoproterozoic ecosystems. Science Advances 1:e1500800. https://doi.org/10.1126/sciadv.1500800

Rau TS, Würfel W, Lenarz T, Majdani O (2013) Three-dimensional histological specimen preparation for accurate imaging and spatial reconstruction of the middle and inner ear. Int J Comput Assist Radiol Surg 8:481–509. https://doi.org/10.1007/s11548-013-0825-7

R Core Team (2016) R: A language and environment for statistical computing. Version 3.3.2. R Foundation for Statistical Computing. https://www.R-project.org/

Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann H (2005) Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. Invest Radiol 40:715–724. https://doi.org/10.1097/01.rli.0000184756.66360.d3

RStudio Team (2016) RStudio: Integrated Development for R. Version 1.0.136. RStudio, Inc. http://www.rstudio.com/

Saalfeld S, Fetter R, Cardona A, Tomancak P (2012) Elastic volume reconstruction from series of ultra-thin microscopy sections. Nat Methods 9:717–720. https://doi.org/10.1038/nmeth.2072

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676–682. https://doi.org/10.1038/nmeth.2019

Schmidbaur H, Keklikoglou K, Faulwetter S, Metscher BD (2015) Exploring methods to remove iodine and phosphotungstic acid stains from zoological specimens. Proc Bruker MicroCT user meeting 2015:116–123

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089

Schulz G, Waschkies C, Pfeiffer F, Zanette I, Weitkamp T, David C, Müller B (2012) Multimodal imaging of human cerebellum – merging X-ray phase microtomography, magnetic resonance microscopy and histology. Sci Rep 2:69. https://doi.org/10.1038/srep00826

Spitzer VM, Whitlock DG (1998) The visible human dataset: the anatomical platform for human simulation. Anat Rec 253:49–57. https://doi.org/10.1002/(SICI)1097-0185(199804)253

Staedler YM, Masson D, Schönenberger J, Sun M (2013) Plant tissues in 3D *via* X-ray tomography: simple contrasting methods allow high resolution imaging. PLoS ONE 8:e75295. https://doi.org/10.1371/journal.pone.0075295

Streicher J, Weninger WJ, Müller GB (1997) External marker-based automatic congruencing: a new method of 3D reconstruction from serial sections. Anat Rec 248:583–602

Toga A, Arnicar T (1985) Image analysis of brain physiology. IEEE Comput Graph Appl 5:20–25. https://doi.org/10.1109/MCG.1985.276259

Turnbull DH, Mori S (2007) MRI in mouse developmental biology. NMR Biomed 20:265–274. https://doi.org/10.1002/nbm.1146

Weninger WJ, Meng S, Streicher J, Müller GB (1998) A new episcopic method for rapid 3-D reconstruction: applications in anatomy and embryology. Anat Embryol 197:341–348. https://doi.org/10.1007/s004290050144

Weninger WJ, Geyer SH, Mohun TJ, Rasskin-Gutman D, Matsui T, Ribeiro I, Costa LdF, Izpisúa-Belmonte JC, Müller GB (2006) High-resolution episcopic microscopy: a rapid technique for high detailed 3D analysis of gene activity in the context of tissue architecture and morphology. Anat Embryol 211:213–221. https://doi.org/10.1007/s00429-005-0073-x

Werneburg I, Wilson LAB, Parr WCH, Joyce WG (2015) Evolution of neck vertebral shape and neck retraction at the transition to modern turtles: an integrated geometric morphometric approach. Syst Biol 64:187–204. https://doi.org/10.1093/sysbio/syu072

Electronic Supplementary Material

[The following Electronic Supplementary Material can be found at the end of this thesis, pp. 266-280]

Online Resource 1 Protocol for paraffin embedding of *Alytes obstetricans* (ZMH A12442)

Online Resource 2 Scripts and documentation for Amira® macro *MultiExport*

Online Resource 3 Descriptive statistics of the sets of shortest distances between the observed midpoints and the, respectively, fitted lines

Online Resource 4 Results of Kruskal–Wallis rank sum test and Dunn's test on distance sets

- Chapter six -

Ontogenetic development of the shoulder joint muscles in frogs (Amphibia: Anura) assessed by digital dissection with implications for interspecific muscle homologies and nomenclature

Karolin Engelkes[1], Supak Panpeng[1], Alexander Haas[1]

1) Center of Natural History (CeNak), Universität Hamburg, Germany

Submitted to Zoomorphology.

Contributions of KE:

Study design: 100%

Data collection: Histology ~ 70%, digitization of histological sections ~ 90%, alignment 100%, contrast staining for μCT 100%, segmentation & 3d reconstruction 100%

Data analysis: 100%

Manuscript drafting: 100%

Contributions confirmed by supervisor: _____

Prof. Dr. Alexander Haas

Abstract

Previous myological studies show inconsistencies with regard to the identification and naming of the shoulder joint muscles in frogs and toads (Amphibia: Anura). Those inconsistencies were revealed and resolved by assessing the ontogenetic development, innervation, and adult morphology of selected anuran species representing ancient lineages and two major neobatrachian groups. To do so, digital dissections of volumes acquired by histological serial sectioning, episcopic microtomy, and contrast-enhanced micro-computed tomography scanning were performed and three-dimensional reconstructions were derived. Muscle units crossing the shoulder joint were defined, their ontogenetic development was described, their homology across species was established, and a consistent nomenclature was suggested. The mm. anconaeus, dorsalis scapulae, latissimus dorsi, and the group of scapulohumeralis muscles were ontogenetically derived from the dorsal pre-muscle mass present in all tetrapods. The ventral pre-muscle mass gave rise to the mm. cleidohumeralis, episternohumeralis, supracoracoideus, coracoradialis, subcoracoscapularis, coracobrachialis, and pectoralis. The results indicate that the mm. anconaeus, dorsalis scapulae, latissimus dorsi, coracoradialis, and the portionis sternalis and abdominalis of the m. pectoralis have consistently been recognized and denoted in previous studies, whereas the names for the muscle units commonly denoted as m. coraco-brachialis longus and as parts of the m. deltoideus are misleading with regard to the ontogenetic origin of these muscles. The mm. scapulohumeralis profundus anterior and posterior, although present, have been overlooked in some studies. The mm. cleidohumeralis, supracoracoideus, and coracobrachialis are present with two parts or portions in some species, these portions have previously not always been recognized and assigned correctly.

Keywords: Homology; metamorphosis; myology; muscle nomenclature; ontogeny; synonym

Introduction

The morphologies of the pectoral fins of fishes and of the forelimbs of tetrapods, as well as of the pectoral girdle that connects them to the axial skeleton, has long received attention in order to establish the homology of the anatomical elements across vertebrates and to understand the evolutionary transformations of fins into limbs and of limbs within tetrapods (e.g., Gegenbaur 1865; Rolleston 1869; Romer 1924; Diogo et al. 2016; Molnar et al. 2018). The evolution and resulting homologies of the skeletal elements of the pectoral girdle and forelimbs are comparably well known and supported by the fossil record (e.g., McGonnell 2001; Coates et al. 2002; Shubin et al. 2006; Shubin et al. 2009; Ponomartsev et al. 2017), whereas the related soft tissues have rarely been considered (Soliz et al. 2018; but see, e.g., Rolleston (1869), Romer (1922), and Abdala and Diogo (2010) for attempts to assess forelimb muscle homologies across tetrapods and Diogo et al. (2016) and Molnar et al. (2018) for the reconstruction of muscle evolution across the fin-to-limb transition).

It is striking that extant amphibians are represented by a caudate species in most of the studies investigating the evolution and homology of anatomical elements of the forelimb and pectoral girdle within vertebrates (e.g., Romer 1922; Diogo and Abdala 2007; Molnar et al. 2018; but see Abdala and Diogo (2010) for the inclusion of an anuran species). It should be kept in mind, however, that the independent evolutionary history of the Caudata is as along as the one of the Anura because both taxa are considered to be sister groups (e.g., Jetz and Pyron 2018; also see the discussion on "basal" or "ancestral" species in, e.g., Krell et al. (2004), Omland et al. (2008), and Zachos (2016)). Consequently, it is likely that derived (i.e., apomorphic) character states have evolved within the Caudata if compared to the last common ancestor of the Anura and Caudata. The reconstruction of the evolution of the forelimbs could therefore benefit from the inclusion of the Anura by contributing to the assessment of muscle character states in the last common ancestor of Anura and Caudata.

Having reassessed existing descriptions of the shoulder joint muscles in different anuran species we observed inconsistencies in the identification of muscle units and in the use of muscle names. Ritland (1955), for example, synonymized his 'supracoracoideus superficialis' with the term 'pectoralis portio epicoracoidea' utilized by Gaupp (1896) and thereby implied the homology of the corresponding muscle units. Diogo and Ziermann (2014), in contrast, reported the presence of both, a pars epicoracoidea of the m. pectoralis and a separate m. supracoracoideus, which contradicts the homology assumption by Ritland (1955). Jones (1933) observed the presence of a m. supracoracoideus profundus in, among others, two bufonid species, whereas Bigalke (1927) reported no such muscle in *Bufo bufo* (then *B. vulgaris*) but a pars superficialis of the m. coraco-brachialis brevis that remarkably resembled the m. supracoracoideus profundus in Jones (1933). The m. scapulohumeralis profundus anterior was observed in various anuran species, including representatives of the Ranidae (Tyson 1987). In a different study (Gaupp 1896), however, no such muscle was reported in species of the genus *Rana*. Likewise, the m. scapulohumeralis profundus posterior was observed in various species, including representatives of the Hyloidae and Ranidae (Tyson 1987), but in other studies this muscle was neither included in a list of pectoral girdle muscles in hylid anurans, nor described in *Rana* (Soliz et al. 2018; Gaupp 1896, respectively).

The inconsistencies noted above make the homologization of shoulder joint muscles across anuran species challenging and obstruct the reconstruction of the character states in the last common ancestor of the Anura. The latter, however, would help to establish muscle homologies between the Anura and the Caudata, and to fit both taxa within the larger picture of the evolution of limbs.

The primary aim of the present study was to identify the muscle units occurring in the shoulder joint of anurans and to establish their inter-specific homologies. To do so, the ontogenetic development and innervation of the shoulder joint muscles were assessed in three anuran species representing one ancient lineage (Bombinatoridae: *Bombina orientalis* (Boulenger, 1890)) and the two major neobatrachian groups (Ranoidea: Ranidae: *Rana temporaria* Linnaeus, 1758; Hyloidea: Bufonidae: *Rhinella marina* (Linnaeus, 1758)). In addition, previously published descriptions of the shoulder joint muscles in selected species (various species of *Rana* in Gaupp 1896; *Bufo bufo* (Linnaeus, 1758) in Bigalke 1927; *Ascaphus truei* Stejneger, 1899 in Ritland 1955) were reassessed in order to identify inconsistencies in the use of muscle names and to suggest a consistent nomenclature. The descriptions by Gaupp (1896) were included because they seem to be the most frequently referenced for anuran muscle anatomy; the nomenclature introduced by him, and modified versions of it, presumably are the most widely used (compare, e.g., Bigalke 1927; Jones 1933; Mahendra 1936; Ritland 1955; Burton 1983; Duellman and Trueb 1994; Manzano et al. 2008; Baleeva 2009). The works of Ritland (1955) and Bigalke (1927) were used for comparison as they provide thorough descriptions of the muscles in species belonging to an ancient anuran linage and the Hyloidea, respectively, and because both refer to the nomenclature of Gaupp (1896).

Material and methods

Specimens and usages

A total of 11 larvae (able 1) ranging from Stage 32 to Stage 41 (staging after Gosner 1960) of *Bombina orientalis*, *Rana temporaria*, and *Rhinella marina* were used to investigate the ontogenetic development and innervation of the shoulder joint muscles by histological serial sectioning and three-dimensional (3d) reconstruction. A pre-existing dataset of *Alytes obstetricans* (Laurenti, 1768) (ZMH A12442) acquired by episcopic microtomy (Engelkes et al. 2018) served as source of data for that species but was also modified and transformed into a hypothetical, schematic anatomical 3d model that illustrates all identified muscle units, their spatial relationships, and their innervations. Contrast enhanced micro-computed tomography (μCT) volumes of adult specimens of *Bufo bufo* (ZMH A04660), *Rhinella marina* (ZMH A15443), and *Ascaphus truei* (UF H 80664; downloaded from MorphoSource, Duke University) were examined. Furthermore, histological serial sections from the ZMH museum collection of a late metamorphic stage of *A. truei* (ZMH A09807, Stage 42) were included.

Table 1 Specimens, developmental stage, and methods. Staging after Gosner (1960). UF: Florida Museum of Natural History; ZMH: Zoologisches Museum Hamburg.

Specimen (Collection number)	**Developmental stage**	**Method**	**Remark**
Alytes obstetricans (ZMH A12442)	subadult / adult	Episcopic microtomy	dataset from Engelkes et al. (2018)
Ascaphus truei (UF H 80664)	adult	µCT	iodine stained; volume downloaded from MorphoSource (doi: 10.17602/M2/M22469)
Ascaphus truei (ZMH A09807)	42	Histology	
Bombina orientalis (ZMH A12427)	32	Histology	
Bombina orientalis (ZMH A12429)	35	Histology	
Bombina orientalis (ZMH A12435)	41	Histology	
Bufo bufo (ZMH A04660)	adult	µCT	unstained and iodine stained; unstained scan published in Engelkes et al. (2020, accepted)
Rana temporaria (ZMH A14736)	32-33	Histology	
Rana temporaria (ZMH A14739)	34	Histology	
Rana temporaria (ZMH A14740)	35	Histology	
Rana temporaria (ZMH A12870)	41	Histology	
Rhinella marina (ZMH A14928)	32-33	Histology, 3d reconstruction	
Rhinella marina (ZMH A14930)	34	Histology, 3d reconstruction	
Rhinella marina (ZMH A14933)	37	Histology, 3d reconstruction	
Rhinella marina (ZMH A14937)	41	Histology, 3d reconstruction	
Rhinella marina (ZMH A15443)	adult	µCT	iodine stained

The latest larval stages examined and, if available, adult specimens of each exemplar species were used to reassess previously published anatomical descriptions (Gaupp 1896; Bigalke 1927; Ritland 1955) of those or closely related (Pyron and Wiens 2011) species. Specimens that were sectioned for this study were deposited at the ZMH collection (Table 1).

Histology

Larval specimens selected for histological serial sectioning were decalcified and embedded in either Paraplast Plus® (Leica Biosystems) or Roti®-Plast with DMSO (Carl Roth GmbH + Co. KG). The resulting blocks were sectioned on a rotatory microtome (Microm™ HM 340 E; Microm International GmbH) with slice thicknesses between 6 and 10 µm. Sections were mounted on glass slides and stained according to an Azan staining protocol (modified from Zbären 1966).

The histological sections of the pectoral girdle region were digitized with a digital microscope (Leica DM6000 B; Leica Microsystems GmbH) and edited (adjustment of brightness and contrast, sharpness, and canvas size) in Fiji (based on ImageJ version 1.51n; Schindelin et al. 2012; Schneider et al. 2012). Depending on the quality of the original series of histological sections, the digital image stacks were either rigidly aligned in Amira®

(version 6.0.1; Konrad-Zuse-Zentrum Berlin, FEI Visualization Sciences Group) or aligned in the Fiji plugin TrakEM2 allowing for affine or elastic transformations (Cardona et al. 2012). The resulting aligned volumes were converted to 8-bit grayscale images.

Episcopic microtomy

The generation of the previously published dataset of episcopic images of *Alytes obstetricans* (ZMH A12442) was described in detail in Engelkes et al. (2018); in short, the procedure was as follows: The specimen was decalcified and impregnated with lead ions followed by Paraplast Plus® embedding. Fiducial points were induced into the block to improve the image alignment quality in consecutive processing steps. Sections of 10 µm thickness were cut of the block surface using a rotatory microtome (Microm™ HM 340 E). Every 30 µm an image of the original block surface was taken (camera: Nikon D7200®; macro lens: Nikon AF-S VR Micro-Nikkor® 105 mm 1:2,8G IF-ED; Nikon Corporation). Then, the surface was stained with a sodium sulfide solution (7 %) and a second picture was taken. Images were digitally processed in IrfanView® (version 4.41; Irfan Skiljan, http://www.irfanview.com; conversion NEF to TIF) and Fiji (subtraction of corresponding unstained and stained images, adjustment of brightness and contrast). One fiducial point per corner was used to align the stack of digital images in Amira®. The obtained volume was converted to 8 bit grayscale in Fiji. In defined distance intervals, sections of the specimen were mounted on glass slides, stained using a modified Azan staining protocol, and digitized as above. The digitized sections were rigidly aligned in Amira®.

MicroCT scanning

A µCT scan of the untreated adult *Bufo bufo* specimen (ZMH A04660) was performed using a YXLON FF35 CT (YXLON International GmbH). Subsequently, the specimen was contrast-stained with Lugol's solution (modified from Metscher 2009; concentration: 1 %, staining duration: 7 days, changed twice) and scanned in a YXLON FF20 CT. The unstained and stained scans were registered in Amira®. The adult specimen of *Rhinella marina* (ZMH A15443) was contrast-stained with Lugol's solution (concentration: 1 %, staining duration: 8 days, changed twice) and µCT scanned in a Phoenix v|tome|x L 450 (GE Sensing & Inspection Technologies GmbH). All µCT scans were performed in an ethanol-saturated atmosphere. The volumes were reconstructed from x-ray projections using the software delivered with the respective scanner.

Segmentation, 3d reconstruction, and visualization

All volumes were imported into Amira® and the shoulder joint muscles and contextual skeletal elements (in larvae: only chondrified or ossified parts, no condensations of mesenchymal cells or perichondrium/periosteum) of the right side were digitally dissected (i.e., segmented; *Segmentation Editor* with *Brush* tool). Bone and cartilage were not distinguished in most specimens. The left instead of the right side was segmented in *Bombina orientalis* ZMH A12427, *Bufo bufo* ZMH A04660, and *Rhinella marina* ZMH A15443 due to tissue visibility/damage on the right side. *Bombina orientalis* ZMH A12427 was mirrored to obtain consistent illustrations. In addition, nerves innervating the shoulder joint muscles were also segmented in the Stage 41 and 42 larvae and in the adult specimen of *Alytes obstetricans*. The original aligned digitized histological sections were used for comparison if possible and necessary. Polygon surfaces of the segmented anatomical elements were created and simplified by iterative reduction of polygon count and smoothing in Amira® and the surfaces were exported in obj format; surface generation and export were accelerated using a custom macro (*MultiExport*; see Engelkes et al. (2018) for details).

The simplified polymesh surfaces of the *R. marina* larvae and *A. obstetricans* were imported into MODO® (version 10.1v2; The Foundry) and manually edited (filling of holes, correction of artefacts, smoothing). The surfaces of some nerves showed considerable artifacts (i.e., discontinuity or holes) and needed major manual editing; care was taken to maintain the important properties (i.e., relative position to other anatomical elements and connections to muscles), but the form and thickness may not represent the actual conditions. The surfaces produced from *A. obstetricans* (ZMH A12442) were used to generate a schematic, generalized anuran model of all muscle portions and respective nerves observed across species by manually modifying the surfaces of the anatomical structures present and adding elements not present in that specimen. All surfaces were given descriptive colors (bone: beige; cartilage: blue; skeletal elements without distinguishing bone and cartilage: gray; muscles: various shades of red such that adjacent muscles had different colors; nerves: yellow) prior to image rendering.

The muscle configurations in the adult specimens of *B. bufo* (ZMH A04660) and *R. marina* (ZMH A15443) were illustrated by combining polygon surfaces (skeleton) and volume renders (muscles) in Amira®; this approach allowed for the visualization of the fiber orientation in some muscles. Final figures were arranged and labelled in Adobe® Illustrator® CS6 (version 16.0.3; Adobe® Systems Software).

Muscle homology and nomenclature

We define shoulder joint muscles pragmatically as the set of all muscles that cross the shoulder joint from their respective origin to their point of insertion. Hypotheses on the inter-specific homology (in terms of primary homology; de Pinna 1991) of muscle units were derived by following the criteria for homology introduced by Remane (1952). Two or more muscle units were considered to be homologous if they showed a similar relation to other (homologous) anatomical structures (first criterion of Remane 1952) and could be connected by similar (intermediate) stages during the ontogenetic development (third criterion of Remane 1952; also see Kerr 1955). The first criterion was mainly applied for the relative locations of the muscle attachment sites and the position of the muscle units to one another. The innervation was also considered as it has been proven useful for the identification and homologization of muscles (e.g., Romer 1924; Holliday and Witmer 2007) and because the branching pattern of the major nerves supplying the forelimb muscles seems to be rather conserved across tetrapods (Hirasawa and Kuratani 2018); yet, it should be noted that some studies (e.g., Cunningham 1890; Romer 1922; Haines 1935; Minkoff 1974) questioned the value of nerve supply for determining muscle homologies.

We suggest a nomenclature for the observed shoulder joint muscles and justify the selection of muscle names in the discussion section. In order to avoid confusing the reader with previous muscle terms, for consistency, we already apply those muscle names in the results section that we eventually suggest on the basis of the evidence presented in this work. In general, the muscle names established by Gaupp (1896), Bigalke (1927), or Ritland (1955) were kept if they were consistently applied in the literature and seemed appropriate to reflect the ontogenetic, and thereby possibly also the evolutionary development. Table 2 provides nomenclatural comparison with the mentioned references. Nerve terminology follows Gaupp (1899).

Results

In the *Bombina*, *Rana*, and *Rhinella* specimens examined (Table 1), the skeleton of the pectoral girdle and forelimbs and the muscles of the shoulder joint developed in parallel and accomplished most of their development while the limb was still inside the branchial cavity. Although there were species-specific differences in the timing of the developmental events, the general pattern of skeletogenesis was as follows: The skeletal elements (except for the dermal bones) arose from various centers of chondrification in pre-cartilaginous

Table 2 Hypotheses on shoulder joint muscle homologies across selected species and implied synonyms of muscle names among different authors. Dotted line: muscles not entirely separated (e.g., continuous at origin).

Suggested name		**Ascaphus truei (Ritland 1955)**	**Rana (Gaupp 1896)**	**Bufo bufo (Bigalke 1927)**
m. coracoradialis [cr]		m. coraco-radialis	m. coraco-radialis	m. coraco-radialis
m. supracoracoideus [sup]	portio anterior [supa]	m. supracoracoideus superficialis	m. pectoralis portio epicoracoidea	m. pectoralis portio epicoracoidea
	portio posterior [supp]			m. coracobrachialis brevis pars superficialis
m. coracobrachialis pars ventralis [cbv]		m. supracoracoideus profundus	m. coraco-brachialis brevis pars superficialis	m. coraco-brachialis brevis pars profunda
m. subcoracoscapularis [sub]		m. subcoracoscapularis	m. coraco-brachialis brevis pars profunda	
m. coracobrachialis pars dorsalis [cbd]		m. coracobrachialis		
m. pectoralis portio coracoidea [pc]			m. coraco-brachialis longus	m. coraco-brachialis longus
m. pectoralis portio sternalis [ps]		m. pectoralis caput sternalis	m. pectoralis portio sternalis	m. pectoralis portio sternalis
m. pectoralis portio abdominalis [pa]		m. pectoralis caput abdominalis	m. pectoralis portio abdominalis	m. pectoralis portio abdominalis
m. scapulohumeralis profundus posterior [shpp]		m. scapulohumeralis profundus posterior	[not described]	m. scapulo-humeralis profundus posterior
m. scapulohumeralis profundus anterior [shpa]		m. scapulohumeralis profundus anterior	m. deltoideus pars scapularis	m. scapulo-humeralis profundus anterior
m. scapulohumeralis superficialis [shs]		m. deltoideus [d2, cleidohumeralis]		m. deltoideus pars scapularis
m. cleidohumeralis [clh]	pars superficialis [clhs]	m. deltoideus [d3, cleidohumeralis]		m. deltoideus pars clavicularis
	pars profunda [clhp]		m. deltoideus pars clavicularis	
m. episternohumeralis [eh]		m. deltoideus [d1, episternohumeralis]	m. deltoideus pars episternalis	m. deltoideus pars cleido-humeralis longus
m. dorsalis scapulae [ds]		m. dorsalis scapulae	m. dorsalis scapulae	m. dorsalis scapulae
m. latissimus dorsi [ld]		m. latissimus dorsi	m. latissimus dorsi	m. latissimus dorsi
m. anconaeus caput scapulare [ancs]		m. anconaeus caput scapulare	m. anconaeus caput scapulare	m. anconaeuscaput scapulare
m. cutaneus pectoris[a] [cp]		-	m. cutaneus pectoris[a]	-

[a] The m. cutaneous pectoris was present in *Rana temporaria* only and was ontogenetically derived from the portionis sternalis and abdominalis of the m. pectoralis; the m. cutaneous pectoris, therefore, might as well have been included in the homologization of the two pectoralis portions.

condensations of mesenchymal cells. The cartilaginous anlagen grew and, where appropriate, fused with one another to form a cartilaginous skeletal precursor element. Eventually, the cartilaginous precursors ossified.

Combining the observations of all larval and adult specimens, we discerned 18 distinct muscle units crossing the shoulder joint (Table 3). Some of these units, however, were

Table 3 Muscles of the shoulder joint with origins and insertions.

Muscle (optional portions) [abbreviations in figures]	**Origin**	**Insertion**	**Comment**
m. anconaeus caput scapulare [ancs]	via strong tendon from posterorventral margin of scapula at the edge of glenoid fossa	proximal part of radioulna, via common tendon with other heads of m. anconaeus	
m. cleidohumeralis (partis superficialis and profunda) [ch, chs, chp]	Anterolateral surface of clavicula, might dorsally be expanded onto acromion	ventral margin and adjacent anterior surface of crista ventralis humeri	split into two partis in *Rana temporaria*: pars superficialis (origin: anterior surface of lateral part of clavicula) and pars profunda (origin: acromion directly ventral to and closely associated with m. scapulohumeralis superficialis; insertion: profound to pars superficialis)
m. coracobrachialis (partis ventralis and dorsalis) [cb, cbv, cbd]	posteroventral surface of lateral coracoid half, adjacent paraglenoid cartilage and scapula, continuous with m. subcoracoscapularis	crista ventralis humeri and adjacent posterior surface of humerus; might be closely associated to / continuous with insertion of m. subcoracoscapularis	partly profound to and in some species entirely continuous with m. subcoracoscapularis; split into two partis in *Rana temporaria*: pars ventralis (origin: coracoid) and pars dorsalis (origin: coracoid, paraglenoid cartilage, and scapula)
m. coracoradialis [cr]	ventral surface of pro- and epicoracoid cartilage, continuous with m. supracoracoideus	via long tendon on proximal part of radioulna	profound to m. supracoracoideus
m. dorsalis scapulae [ds]	lateral surface of suprascapula and cleithrum	proximal part of anterior surface of crista ventralis, via common tendon with m. latissimus dorsi	
m. episternohumeralis [eh]	episternum (if present) or anteromedial part of procoracoid cartilage	posterior surface of humerus, continuous with m. scaphulohumeralis superficialis	
m. latissimus dorsi [ld]	fascia dorsalis or processus transversus of vertebra IV	proximal part of anterior surface of crista ventralis humeri, via common tendon with m. dorsalis scapulae	
m. pectoralis portio abdominalis [pa]	ventral surface of m. rectus abdominis (rectus sheath and tendinous inscriptions)	connective tissue that covers tendon of m. coracoradialis on posterior surface of crista ventralis humeri, distal and/or ventral to insertions of mm. supracoracoideus and pectoralis portio sternalis	
m. pectoralis portio coracoidea	medial part of posterior and ventral surface of coracoid, might be expanded onto epicoracoid cartilage	posterior surface of humerus or crista ventralis humeri, distal to insertion of other portions of m. pectoralis or/and m. coracobrachialis	mostly profound to m. pectoralis portio sternalis

Table 3 continued.

Muscle (optional portions) [abbreviations in figures]	Origin	Insertion	Comment
m. pectoralis portio sternalis [ps]	lateral and ventral surface of the sternum	connective tissue that covers tendon of m. coracoradialis on posterior surface of crista ventralis humeri, proximal and/or dorsal to insertion of m. pectoralis portio abdominalis	mostly superficial to m. pectoralis portio coracoidea
m. scapulohumeralis profundus anterior [shpa]	anterolateral surface of scapula, continuous with origin of scapulohumeralis superficialis (except for *Rana temporaria*)	proximal on anterior base of the crista ventralis humeris; proximal to common tendon of mm. dorsalis scapulae and latissimus dorsi	profound to m. scapulohumeralis superficialis
m. scapulohumeralis profundus posterior [shpp]	ventral part of posterolateral surface of scapula	anterodorsal surface of proximal part of humerus	
m. scapulohumeralis superficialis [shs]	anterolater surface of scapula, might be ventrally expanded onto acromion, continuous with origin of m. scapulohumeralis profundus anterior; in *Rana temproaria*: medial surface of pars glenoidalis of scapula, adjacent cartilage, and lateral part of coracoid, not continuous with origin of m. scapulohumeralis profundus anterior	Posterior to ventral surface of humerus, continuous with m. episternohumeralis	superficial to m. scapulohumeralis profundus anterior
m. subcoracoscapularis [sub]	posteroventral part of medial surface of scapula and adjacent paraglenoid cartilage; continuous with m. coracobrachialis	posterodorsal part of distal humerus or posterior humerus surface and/or connective tissue covering the tendon of m. coracoradialis, distal to insertion of m. coracobrachialis; in *Rana temporaria*: more proximal, mostly continuous with insertion of m. coracobrachialis pars dorsalis	superficial and in *Rana temporaria* almost entirely fused to m. coracobrachialis
m. supracoracoideus (anterior and posterior) [sup, supa, supp]	ventral surface of pro- and epicoracoid cartilage and adjacent part of coracoid; continuous with m. coracoradialis	connective tissue that covers tendon of m. coracoradialis on posterior surface of crista ventralis humeri, proximal to insertion of m. pectoralis	superficial to m. coracoradialis; in *Rhinella marina* and *Bufo bufo* split into two portions: portio anterior (origin: pro- and epicoracoid cartilage) and portio posterior (origin: coracoid)

species-specifically fused or did not separate during ontogeny. The following general pattern of myogenesis was found in the ontogenetic series examined: The shoulder joint muscles differentiated from condensations of pre-muscle cells. In the earliest developmental stages, Stages 32-33, the anlage of the m. anconaeus and three pre-muscle masses were present; one of the pre-muscle mass was located dorsal, the other two ventral to the humerus. In subsequent stages, the condensations of the pre-muscle cells split into smaller units that differentiated into the distinct shoulder joint muscles. The pre-muscle mass located dorsal to the humerus differentiated into the mm. dorsalis scapulae, latissimus dorsi, scapulohumeralis superficialis, scapulohumeralis profundus anterior, and scapulohumeralis profundus posterior. The anterior mass of the two pre-muscle masses located ventral to the humerus split into the mm. cleidohumeralis, episternohumeralis, coracoradialis, and supracoracoideus, whereas the posteroventral pre-muscle mass differentiated into the mm. pectoralis, coracobrachialis, and subcoracoscapularis; this posteroventral pre-muscle mass also gave rise to the m. cutaneus pectoris in *Rana temporaria*. The two pre-muscle masses ventral to the developing humerus were barely separable in the earliest developmental stage of *R. temporaria* (ZMH A14736). Nerves were present and in contact with the pre-muscle masses or muscle units in all stages examined, even the earliest ones.

Species-specific muscle variations

Eighteen distinct muscle units crossing the shoulder joint were observed in the species examined (Figs 1-3, Table 3), but not all these units were present in all specimens. The muscles and their respective origins and insertions are described in Table 3; in the following, only interspecific differences are reported.

In *Ascaphus truei*, the mm. scapulohumeralis superficialis, cleidohumeralis, and episternohumeralis formed one continuous muscle complex. The anteromedial part of the m. supracoracoideus was continuous with this muscle complex, but distinct from it at the insertion. A large part of the m. supracoracoideus was covered by the anteriorly expanded portio sternalis of m. pectoralis. The m. coracobrachialis was split into a pars dorsalis and a pars ventralis; the pars dorsalis was continuous with the portio coracoidea of the m. pectoralis. The m. coracobrachialis and the portio coracoidea of the m. pectoralis were continuous at their insertions in *Alytes obstetricans*; the insertions of these two muscles were adjacent in the latest larval stage considered (Stage 41) of *Bombina orientalis*.

Rhinella marina and *Bufo bufo* were very similar with respect to the configuration of their shoulder joint muscles (Supplementary Material Figs S1-S3). The major difference

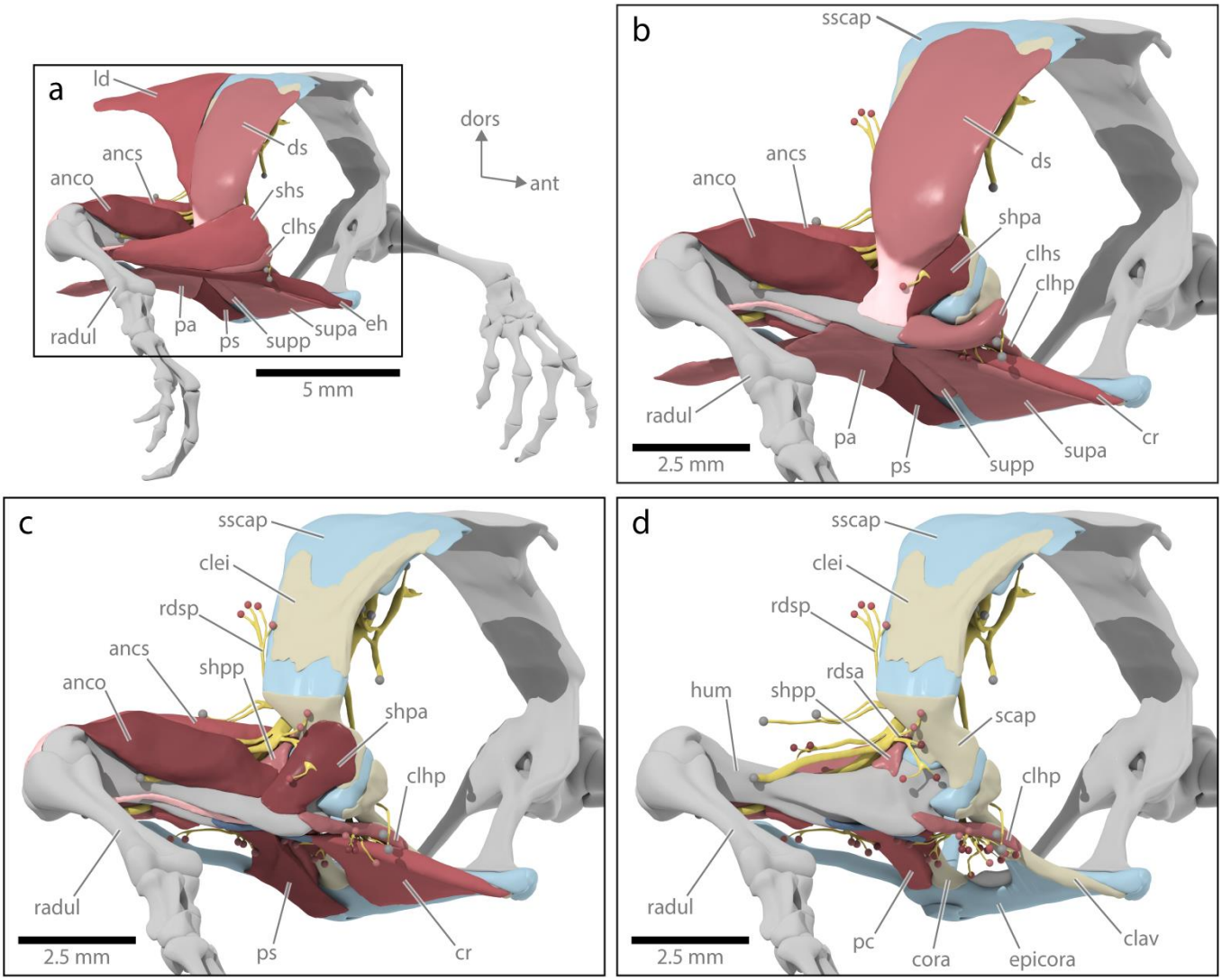


Fig. 1 Hypothetical generalized pattern of shoulder joint muscles with respective nerve supplies in anurans. **a–d** Anterolateral views, muscle layers successively removed. Surface model originally derived from a 3d representation of *Alytes obstetricans* (ZMH A12442) to serve as a model, but manually modified (mm. cleidohumeralis and supracoracoideus split into two portions each, nerve supplies adjusted) and no longer representing the character states of that species. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Red: muscles (different shades for better visual separation of adjacent muscles); yellow: nerves; beige: bone; light blue: cartilage; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; ant: anterior; clav: clavicula; clei: cleithrum; clhp: m. cleidohumeralis pars profunda; clhs: m. cleidohumeralis pars superficialis; cora: coracoid; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; ld: m. latissimus dorsi; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; radul: radioulna; rdsa: r. dorsalis scapulae anterior; rdsp: r. dorsalis scapulae posterior; scap: scapula; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sscap: suprascapula; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior.

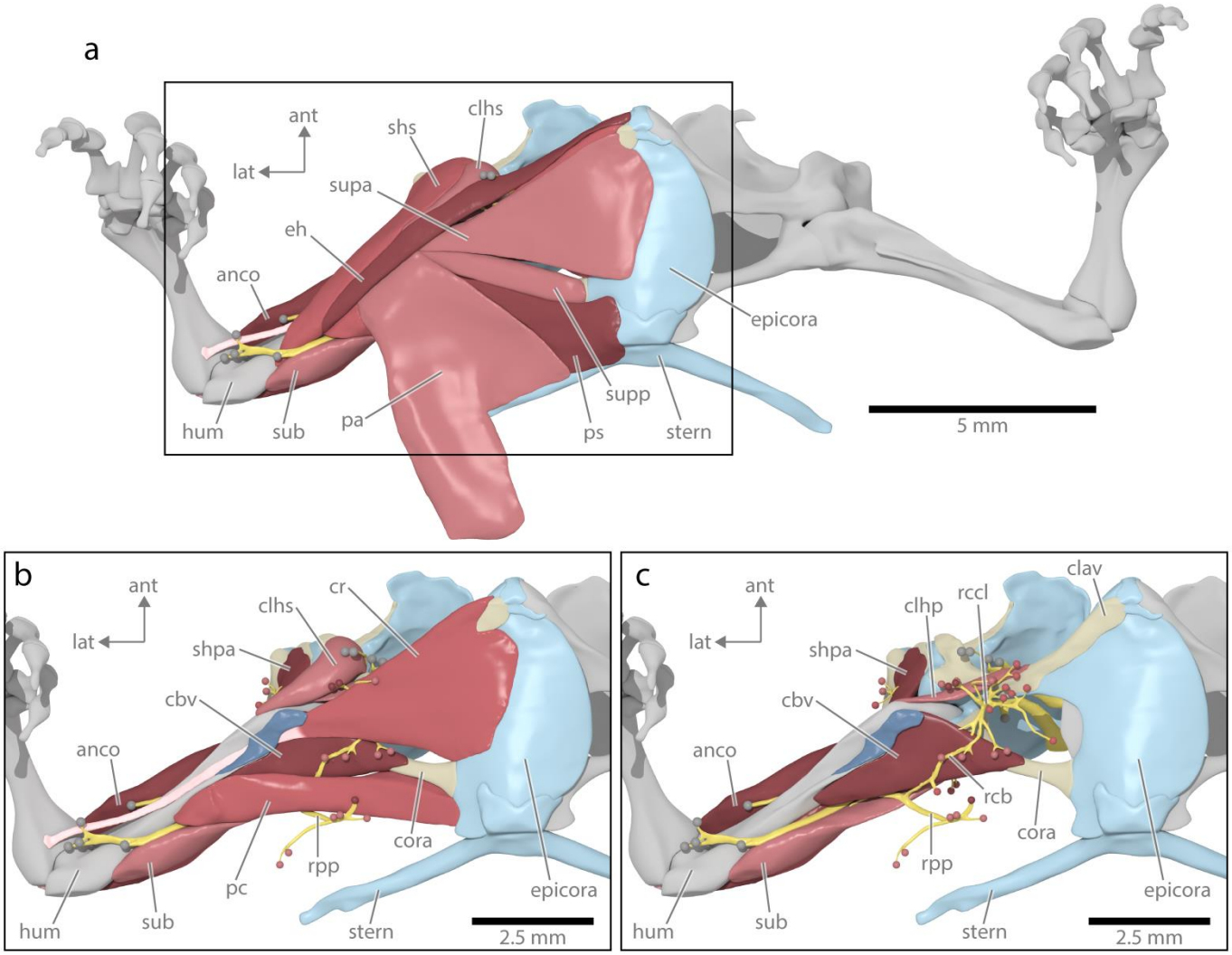


Fig. 2 Hypothetical generalized pattern of shoulder joint muscles with respective nerve supplies in anurans. **a–c** Ventral views, muscle layers successively removed. Same surface model as in Fig. 1. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Red: muscles (different shades for better visual separation of adjacent muscles); yellow: nerves; beige: bone; light blue: cartilage; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. anco: heads of m. anconaeus not crossing the shoulder joint; ant: anterior; cbv: m. coracobrachialis pars ventralis; clav: clavicula; clhp: m. cleidohumeralis pars profunda; clhs: m. cleidohumeralis pars superficialis; cora: coracoid; cr: m. coracoradialis; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; rcb: r. coraco-brachialis; rccl: r. coraco-clavicularis; rpp: r. pectoralis proprius; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; stern: sternum; sub: m. subcoracoscapularis; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior.

between these species concerned the m. episternohumeralis: In *R. marina*, this muscle was continuous with the m. supracoracoideus at its origin and along most of its length (muscles artificially separated for illustrational porous), whereas in *B. bufo*, these muscles were clearly separated at their origins and along most of their lengths. In *B. bufo*, there was a tendency

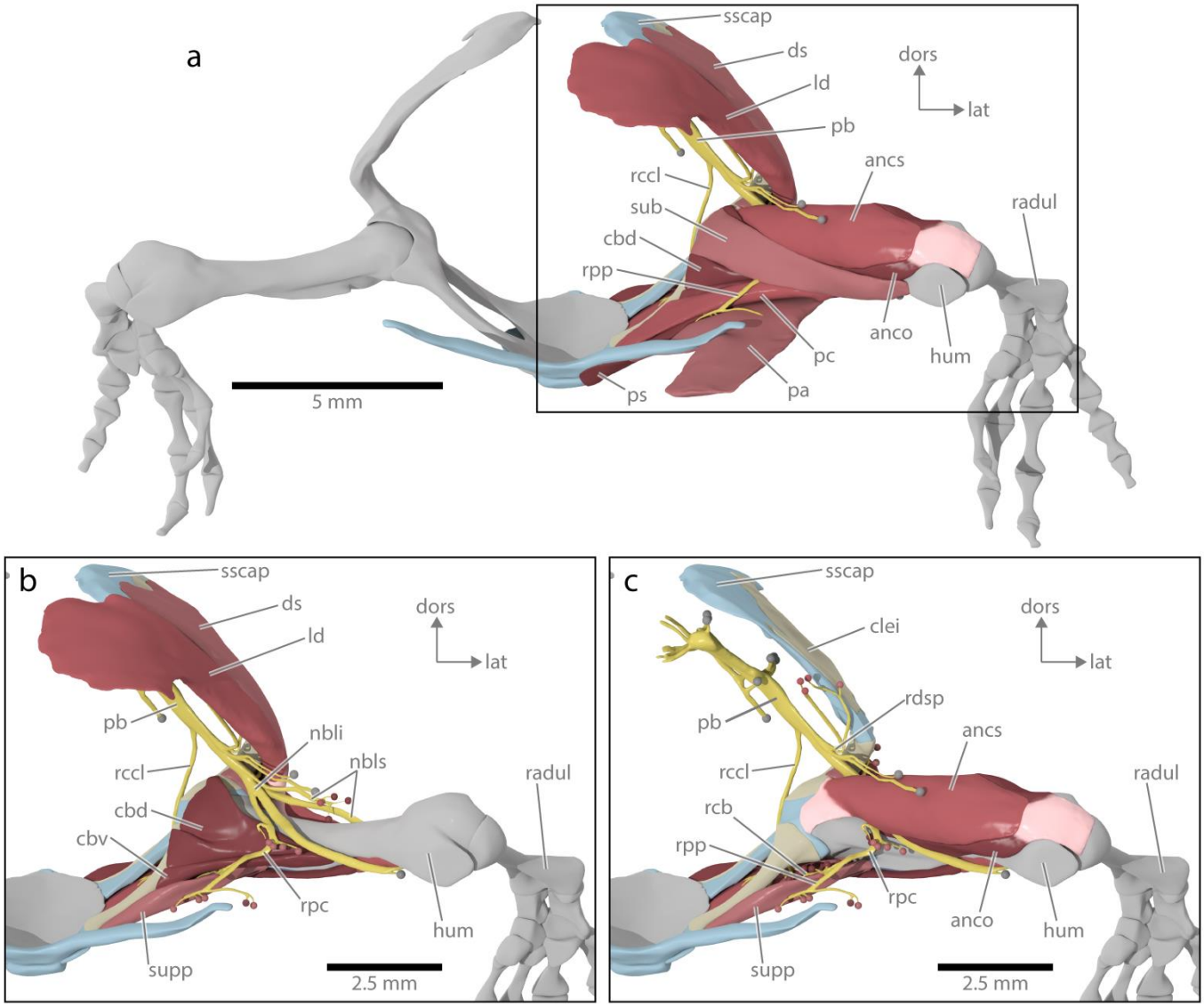


Fig. 3 Hypothetical generalized pattern of shoulder joint muscles with respective nerve supplies in anurans. **a–c** Posterior views, muscle layers successively removed. Same surface model as in Fig. 1. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Red: muscles (different shades for better visual separation of adjacent muscles); yellow: nerves; beige: bone; light blue: cartilage; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; cbd: m. coracobrachialis pars dorsalis; cbv: m. coracobrachialis pars ventralis; dors: dorsal; ds: m. dorsalis scapulae; hum: humerus; ld: m. latissimus dorsi; lat: lateral; nbls: n. brachialis longus superior; nbli: n. brachialis longus inferior; pa: m. pectoralis portio abdominalis; pb: plexus brachialis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; radul: radioulna; rcb: r. coraco-brachialis; rccl: r. coraco-clavicularis; rpc: r. pectoralis communis; rpp: r. pectoralis proprius; sub: m. subcoracoscapularis; supp: m. supracoracoideus portio posterior.

towards the formation of a pars ventralis and a pars dorsalis within the m. coracobrachialis, but these parts were mostly continuous and may not be considered to be distinct muscle units. In both species, the m. supracoracoideus was present with an anterior and a posterior muscle

unit; these two parts together claimed approximately the same space on the ventral part of the girdle as the undivided m. supracoracoideus in the other species.

In the latest developmental stage of *Rana temporaria* considered herein (Stage 41), the m. scapulohumeralis superficialis originated from the medial surface of the pars glenoidalis of the scapula, the adjacent cartilage, and the lateral part of coracoid; the origin of this muscle was not continuous with the origin of the m. scapulohumeralis profundus anterior. The mm. cleidohumeralis and coracobrachialis were each present with two parts that were continuous at their insertions. In addition, the pars superficialis of the m. cleidohumeralis was barely separable form the m. scapulohumeralis superficialis at its origin. Both parts of the m. coracobrachialis together claimed about the same space as the undivided m. coracobrachialis in the other species. The m. subcoracoscapularis was closely associated and mostly continuous with the pars dorsalis of the m. coracobrachialis; the former inserted more proximal on the humerus than observed in the other species. A m. cutaneus pectoris was present and originated from the ventral surface of m. rectus abdominis and inserted onto the skin.

Species-specific innervation of the shoulder joint muscles

The shoulder joint muscles were innervated by various branches of the plexus brachialis that was formed by fibers of the nn. spinalis II-IV (Fig. 4). The rr. dorsalis scapulae anterior and posterior arose from about the lateral aspect of the plexus brachialis. The r. dorsalis scapulae anterior innervated the m. dorsalis scapulae and the group of scapulohumeralis muscles. It either passed through the m. scapulo humeralis profundus anterior before it ended in the m. scapulohumeralis superficialis (*Alytes obstetricans*, *Rana temporaria*; Fig. 1c) or entered between these two muscles (*Ascaphus truei*, *Bombina orientalis*, *Rhinella marina*; Fig. 5a, b). The r. dorsalis scapulae posterior innervated the posterior part of the m. dorsalis scapulae and the m. latissimus dorsi.

The r. coraco-clavicularis arose from the anterior aspect of the plexus brachialis, passed through the opening between the procoracoid and coracoid, and innervated the mm. coracoradialis, supracoracoideus (only anterior part or portion), episternohumeralis, and cleidohumeralis (both portions if two were present). In all species examined, the ramus innervating the m. coracoradialis was the first to separate from the r. coraco-clavicularis. In *Ascaphus truei*, *B. orientalis*, and *Alytes obstetricans*, there was a separate branch arising from the r. coraco-clavicularis that formed an anastomosis with the r. coraco-brachialis of the r. pectoralis communis of the n. brachialis longus inferior; the anastomosis and the respective

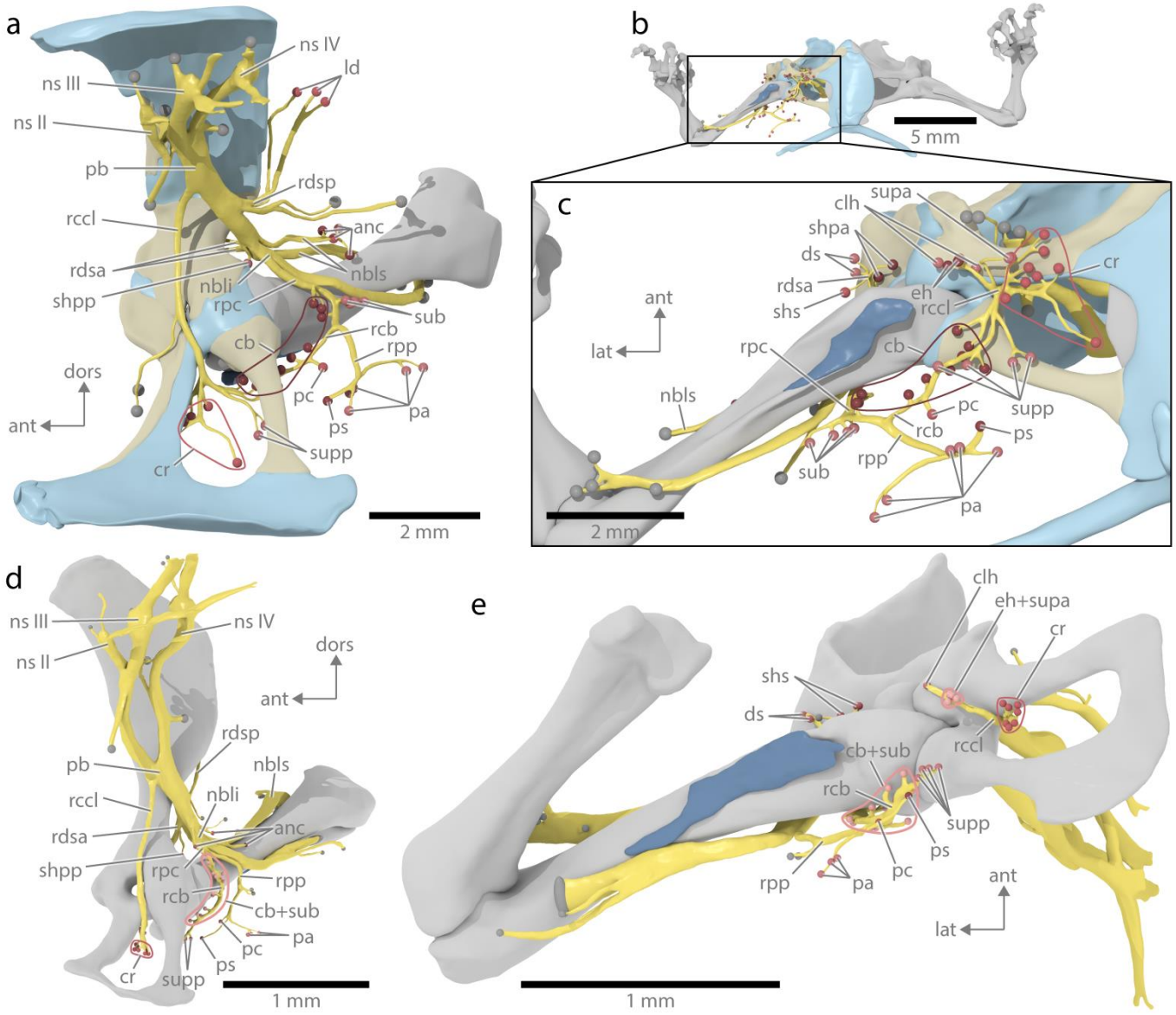


Fig. 4 Plexus brachialis and associated nerve branches innervating the shoulder joint muscles. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Yellow: nerves; beige: bone; light blue: cartilage; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. **a–c** Generalized pattern in anurans; same surface model as in Fig. 1. **a** medial view of right pectoral girdle half and humerus. **b** Ventral view and **c** detail of ventral view of pectoral girdle and forelimbs. **d–e** Pattern of shoulder joint muscle innervation in a *Rhinella marina* larva (Stage 41, ZMH A14937). **d** Medial view of right pectoral girdle half and humerus, anterior to the left. **e** Ventral view of right pectoral girdle half and humerus, anterior to the top. anc: m. anconaeus; ant: anterior, cb: m. coracobrachialis; clh: m. cleidohumeralis; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; lat: lateral; ld: m. latissimus dorsi; nbli: n. brachialis longus inferior; nbls: n. brachialis longus superior; ns: n. spinalis; pa: m. pectoralis portio abdominalis; pb: plexus brachialis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; rccl: r. coraco-clavicularis; rcb: r. coraco-brachialis; rdsa: r. dorsalis scapulae anterior; rdsp: r. dorsalis scapulae posterior; rpc: r. pectoralis communis; rpp: r. pectoralis proprius; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sub: m. subcoracoscapularis; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior.

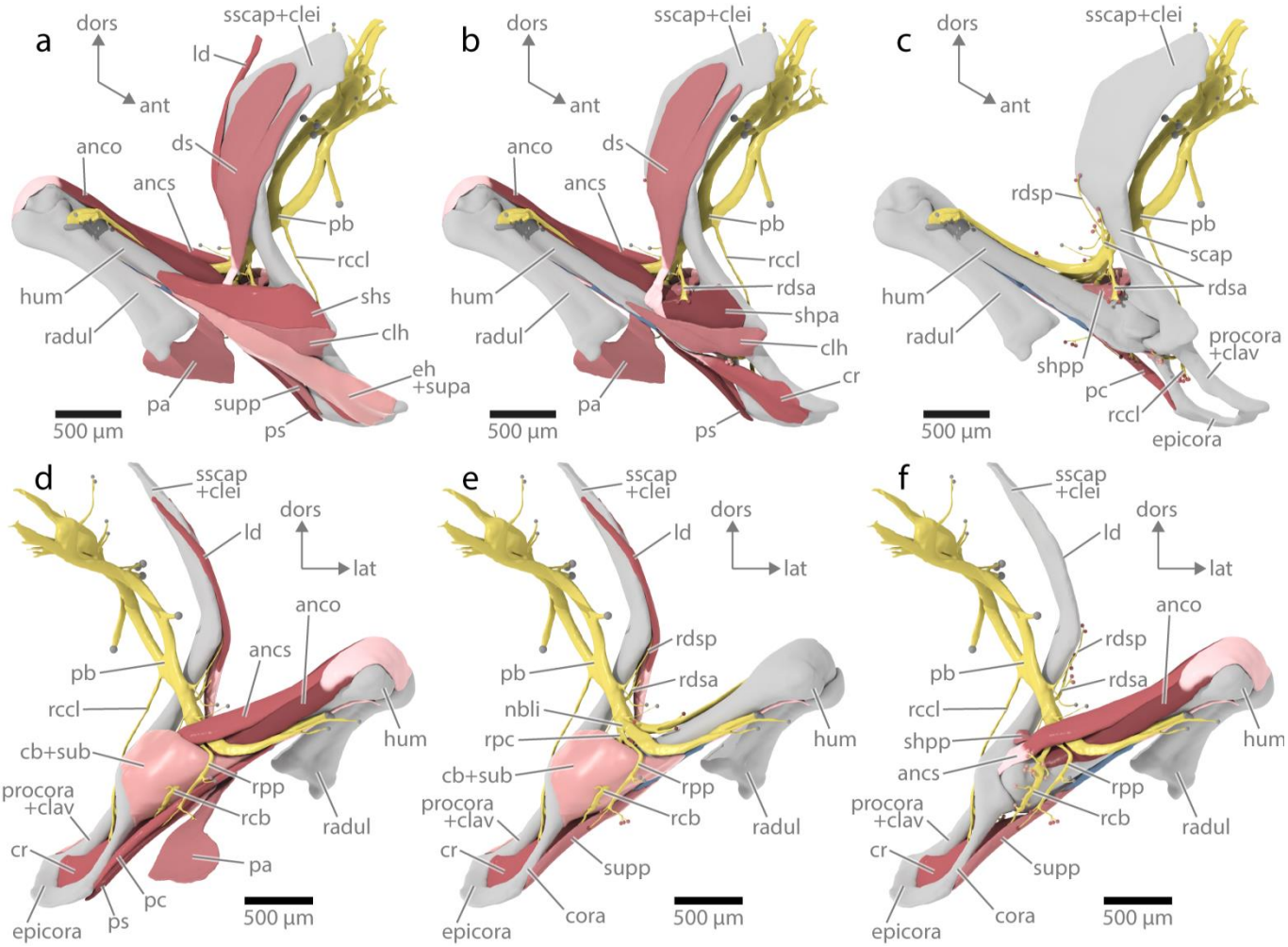


Fig. 5 Shoulder joint muscles with respective nerve supplies in *Rhinella marian* (Stage 41, ZMH A14937). Surfaces derived from aligned histological serial sections, muscle layers successively removed. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Red: muscles (different shades for better visual separation of adjacent muscles); yellow: nerves; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. **a–c** Anterolateral views of right pectoral girdle half. **d–f** Posterior views of right pectoral girdle half. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; ant: anterior; cb: m. coracobrachialis; clav: clavicula; clei: cleithrum; clh: m. cleidohumeralis; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; ld: m. latissimus dorsi; nbli: n. brachialis longus inferior; pa: m. pectoralis portio abdominalis; pb: plexus brachialis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; radul: radioulna; rccl: r. coraco-clavicularis; rcb: r. coraco-brachialis; rdsa: r. dorsalis scapulae anterior; rdsp: r. dorsalis scapulae posterior; rpc: r. pectoralis communis; rpp: r. pectoralis proprius; scap: scapula; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sscap: suprascapula; sub: m. subcoracoscapularis; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior.

rami were located dorsal (profound) to the mm. coracoradialis, supracoracoideus, and the portions of the m. pectoralis, whereas the laid ventral (superficial) to the m. coracobrachialis (Fig. 2). No such anastomosis was present in *R. marina* (Fig. 6) or *Rana temporaria*.

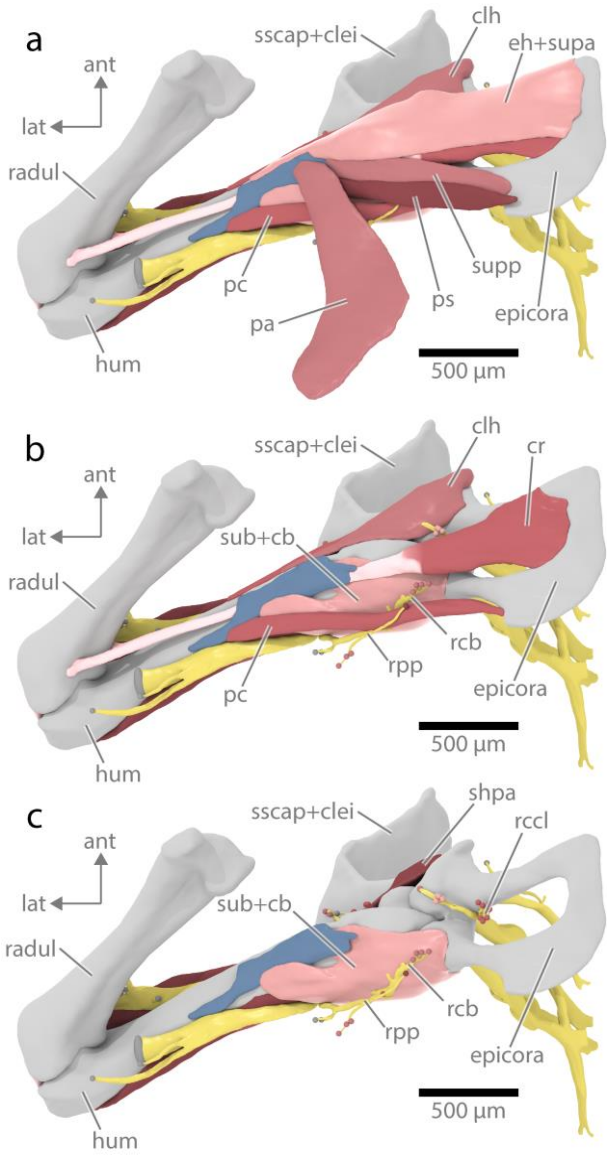


Fig. 6 Shoulder joint muscles with respective nerve supplies in *Rhinella marina* (Stage 41, ZMH A14937). Ventral views, muscle layers successively removed; mm. dorsalis scapulae and latissimus dorsi not shown. Surfaces derived from aligned histological serial sections. Spheres: nerve ending in muscle (red spheres) or cut (grey, no connection to shoulder joint muscles). Red: muscles (different shades for better visual separation of adjacent muscles); yellow: nerves; dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. ant: anterior, cb: m. coracobrachialis; clei: cleithrum; clh: m. cleidohumeralis; cr: m. coracoradialis; ds: m. dorsalis scapulae; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; radul: radioulna; rccl: r. coraco-clavicularis; rcb: r. coraco-brachialis; rpp: r. pectoralis proprius; shpa: m. scapulohumeralis profundus anterior; sscap: suprascapula; sub: m. subcoracoscapularis; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior.

The n. brachialis longus inferior separated from the posterior aspect of the plexus brachialis and gave rise to, among others, the r. pectoralis communis, which in turn split into the rr. coraco-brachialis and pectoralis proprius. The r. pectoralis communis and its derivatives laid profound to the caput scapulare of the m. anconaeus. In *Ascaphus truei*, *B. orientalis*, and *Alytes obstetricans*, the r. pectoralis communis passed between the mm. subcoracoscapularis and coracobrachialis (Fig. 3a, b), whereas the rr. coraco-brachialis and pectoralis proprius separated before they entered the space between the mm. subcoracoscapularis and coracobrachialis in *Rhinella marina* and *Rana temporaria* (Fig. 5d, e). In the latter two species, only the r. coraco-brachialis passed between the two muscles, the r. pectoralis proprius laid superficial to them. In all species examined, the r. coraco-brachialis innervated the mm. subcoracoscapularis, coracobrachialis, and the posterior part or portion of the m. supracoracoideus. In *A. obstetricans* (Fig. 4a, b) and *R. temporaria*, the r. coraco-

brachialis also innervated the portio coracoidea of the m. pectoralis, whereas this portion was innervated by the r. pectoralis proprius in *B. orientalis* and *Rhinella marina* (Figs 4d, f, 5d, e). No distinct portio coracoidea was observed in *Ascaphus truei*, but the muscle fibers most likely representing the portio coracoidea (as inferred from the relative position to the rr. coraco-brachialis and pectoralis proprius) received nerve supply from both, the rr. coraco-brachialis and pectoralis proprius. The r. pectoralis proprius of the r. pectoralis communis innervated the portionis sternalis and abdominalis of the m. pectoralis in all species examined and, as described above, in *B. orientalis*, *R. marina*, and *A. truei* also the portio coracoidea. In *Rana temporaria*, the r. pectoralis proprius also supplied the m. cutaneus pectoris.

The n. brachialis longus superior separated from the posterior aspect of the plexus brachialis in close proximity to the n. brachialis longus inferior. It innervated, among others, the heads of the m. anconaeus with one exception: In *Rhinella marina*, the caput scapulare received nerve supply from two rami that arose from the plexus brachialis in close proximity to the base of the n. brachialis longus superior.

Development of the skeleton

Precursors of the pectoral girdle skeleton and long bones of the forelimbs were present and developed throughout all larval stages considered herein. Despite species-specific differences, the following general pattern was observed: In the earliest of the considered stages, the pectoral girdle skeleton was present by cartilaginous precursors of the elements (scapula and coracoid) that form the glenoid fossa and, case-specifically, by the ventral parts of the suprascapula or condensations of cells preceding the suprascapula. There case-specifically also was a separate precursors of the procoracoid cartilage. During the course of development, the coracoid and procoracoid extended ventrally and, in later stages, grew towards one another (formation of epicoracoid) with cartilaginous tissue.

With about the onset of the climax of the metamorphosis, the orientation of the skeletal complex formed by the coracoid, procoracoid, and epicoracoid stated to shift from a more vertical towards a rather horizontal orientation. In addition, these ventral elements grew towards their counterparts of the other girdle half. The scapula and suprascapula extended dorsally. Humerus, radius, and ulna lengthened throughout all developmental stages and approached their adult form (i.e., development of crests, fusion of radius and ulna). The ossification of the endochondral bones, as well as the development of the dermal bones (clavicula and cleithrum), began in bone- and species-specific stages. More detailed descriptions of the chondrogenesis and ossification during larval development and

metamorphosis have been published elsewhere (e.g., Púgener and Maglia 1997; Maglia and Púgener 1998; Baleeva 2001, 2009; Shearman 2005, 2008; Havelková and Roček 2006) and are beyond the scope of this study.

Development of the shoulder joint muscles in *Rhinella marina*

In the earliest larval stage of *Rhinella marina* considered herein (Stage 32-33; ZMH A14928; Fig. 7a, b), the muscles dorsal the shoulder joint were present by a distinct caput scapulare of the m. anconaeus (distally continuous with other heads of m. anconaeus), a common precursor of the mm. dorsalis scapulae and latissimus dorsi, one pre-muscle masses representing the mm. scapulohumeralis profundus anterior and superficialis, as well as a distinct mass representing the m. scapulohumeralis profundus posterior. Both precursors of the scapulohumeralis muscles were connected by a loose accumulation of undifferentiated cells. Ventrally, one pre-muscle mass preceded the mm. supracoracoideus (portionis anterior and posterior), episternohumeralis, and cleidohumeralis. This mass was located superficial to and was medioventrally continuous with the precursor of m. coracoradialis. Posteroventrally, the mm. subcoracoscapularis, coracobrachialis, and all portions of m. pectoralis were present by one common pre-muscle mass; the cells representing the future m. pectoralis seemed less differentiated than the cells preceding the other muscles, so the m. pectoralis could artificially be separated from the others.

In Stage 34 (ZMH A14930; Fig. 7c, d), the undifferentiated cells connecting the m. scapulohumeralis profundus posterior to the mm. scapulohumeralis profundus anterior and superficialis disappeared and the m. latissimus dorsi became distinguishable from the m. dorsalis scapulae. Anteroventrally, the m. cleidohumeralis, as well as the portio posterior of the m. supracoracoideus split from the remaining pre-muscle mass representing the future portio anterior of the m. supracoracoideus and the m. episternohumeralis; the separations were most obvious at the origins, whereas the muscles were almost continuous at their insertions. The posteroventral muscle precursor split into several (pre-)muscle masses: The m. pectoralis portio abdominalis became entirely distinct, while the portionis sternalis and coracoidea were separated at their insertions (the insertion of the portio coracoidea shifted to a more distal position), but remained continuous at the origin. The mm. coracobrachialis and subcoracoscapularis formed one separate muscle mass.

The specimens of Stages 37 (ZMH A14933; Fig. 7e-h) and 41 (ZMH A14937; Figs 5, 6) showed all muscle units also present in the adult specimen (ZMH A15443; Supplementary Material Figs S1-S3): The mm. scapulohumeralis superficialis and profundus anterior

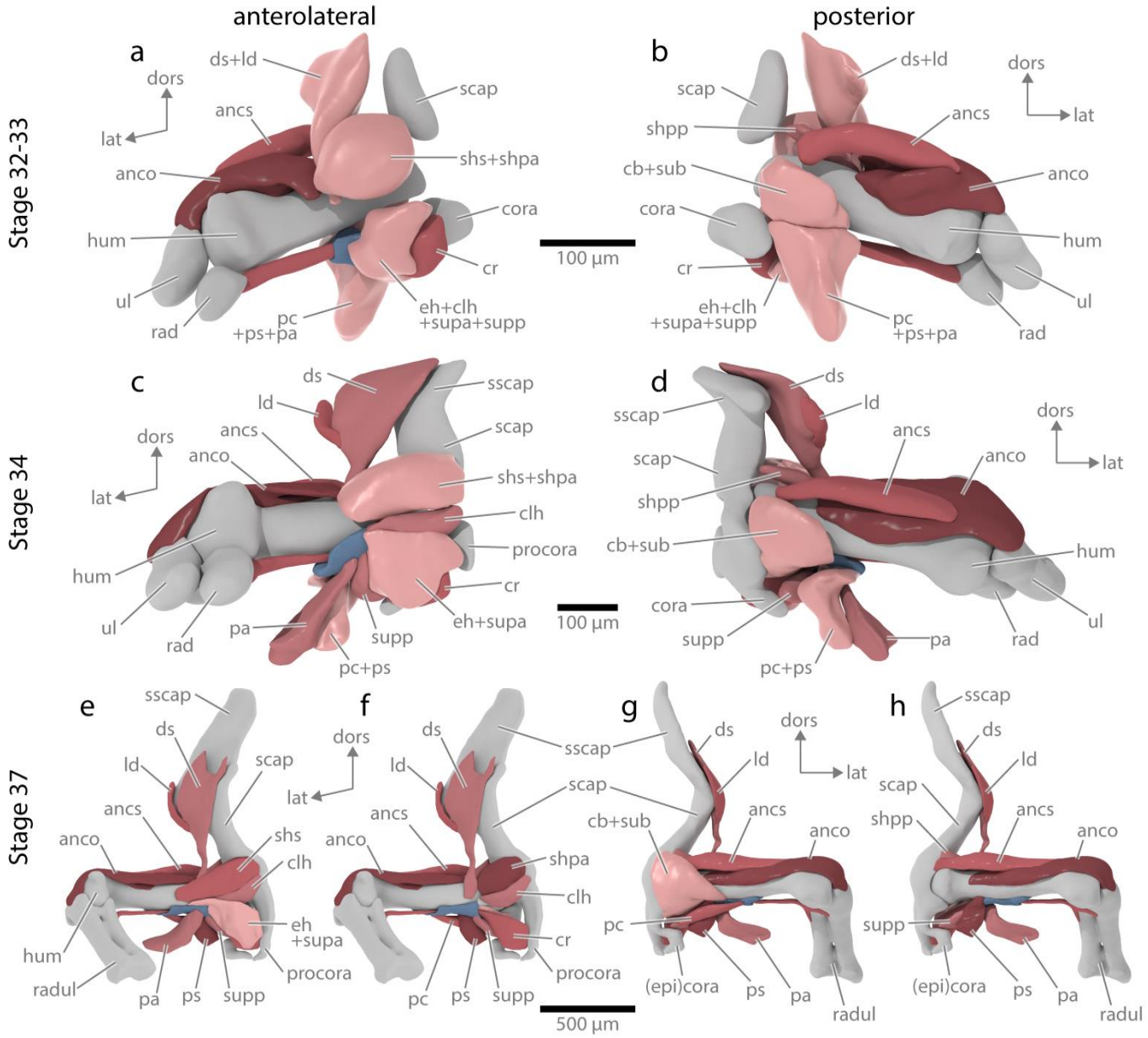


Fig. 7 Ontogenetic development of the shoulder joint muscles in *Rhinella marina*. Anterolateral (left) and posterior (right) views. Surfaces derived from aligned histological serial sections. Red: muscles (different shades for better visual separation of adjacent muscles); dark blue: connective tissue; light gray: skeletal element with no distinction of bone and cartilage. **a–b** Stage 32-33 larva (ZMH A14928); separation if cb+sub and pc+ps+pa artificial based on differences in the cell differentiation. **c–d** Stage 34 larva (ZMH A14930). **f–h** Stage 37 larva (ZMH A14933), muscle layers successively removed. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; cb: m. coracobrachialis; clh: m. cleidohumeralis; cora: coracoid; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; ld: m. latissimus dorsi; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; rad: radius; radul: radioulna; scap: scapula; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sscap: suprascapula; sub: m. subcoracoscapularis; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior; ul: ulna.

were continuous at their origins, but clearly separated at their insertions. The same was observed for the m. episternohumeralis and the portio anterior of the supracoracoideus, as well as for mm. coracobrachialis and subcoracoscapularis. At their origins, the muscles of these complexes could only artificially be separated in the adult specimen by tracing the muscles fibers from the insertion to the origin. The portionis sternalis and coracoidea of the m. pectoralis were entirely separated.

Development of the shoulder joint muscles in *Bombina orientalis*

The overall pattern of the development of the shoulder joint muscles in *Bombina orientalis* was similar to the one observed in *Rhinella marina*, but there were some differences. Most strikingly, most developmental events occurred in earlier larval stages than in *R. marina*. Other than in *R. marina* (m. episternohumeralis continuous with m. supracoracoideus in Stage 32-33), the mm. episternohumeralis and cleidohumeralis formed one independent pre-muscle mass in the earliest considered developmental stage of *B. orientalis* (Stage 32, ZMH A12427); two portions were recognizable within this pre-muscle mass based on the cell orientation (Fig. 8a). As in later developmental stages of *R. marina*, the mm. scapulohumeralis superficialis and profundus anterior in *B. orientalis* (ZMH A12427) were continuous at their origins but clearly separated at their insertions (Fig. 8b). Both these muscles had no connection to the m. scapulohumeralis profundus posterior. The mm. dorsalis scapulae and latissimus dorsi were distinct from each other as well. The caput scapulare of the m. anconaeus was distinct at its origin and distally continuous with the other heads of the m. anconaeus. Posteriorly, the precursors of the portionis abdominalis and sternalis of the m. pectoralis were mostly continuous, but distinct at their insertions; the portio coracoidea was entirely separated from the former two portions. The mm. subcoracoscapularis and coracobrachialis were continuous at their origins and separated at the insertions (Fig. 8c).

In the *B. orientalis* specimen of Stage 35 (ZMH A12429), the mm. episternohumeralis and cleidohumeralis remained mostly continuous at their insertions but were clearly separated at the origins (Fig. 8d). Within the m. supracoracoideus there was a tendency towards forming an anterior and a posterior portion (Fig. 8e), but the portions were closely associated and could not be considered distinct. The portionis abdominalis and sternalis of the m. pectoralis were distinct from one another (Fig. 8f).

The muscle configuration in Stage 41 (ZMH A12435) resembled the one observed in Stage 35 with the exceptions that the mm. episternohumeralis and cleidohumeralis, although laterally in close proximity to each other, were separated at their origins and insertions

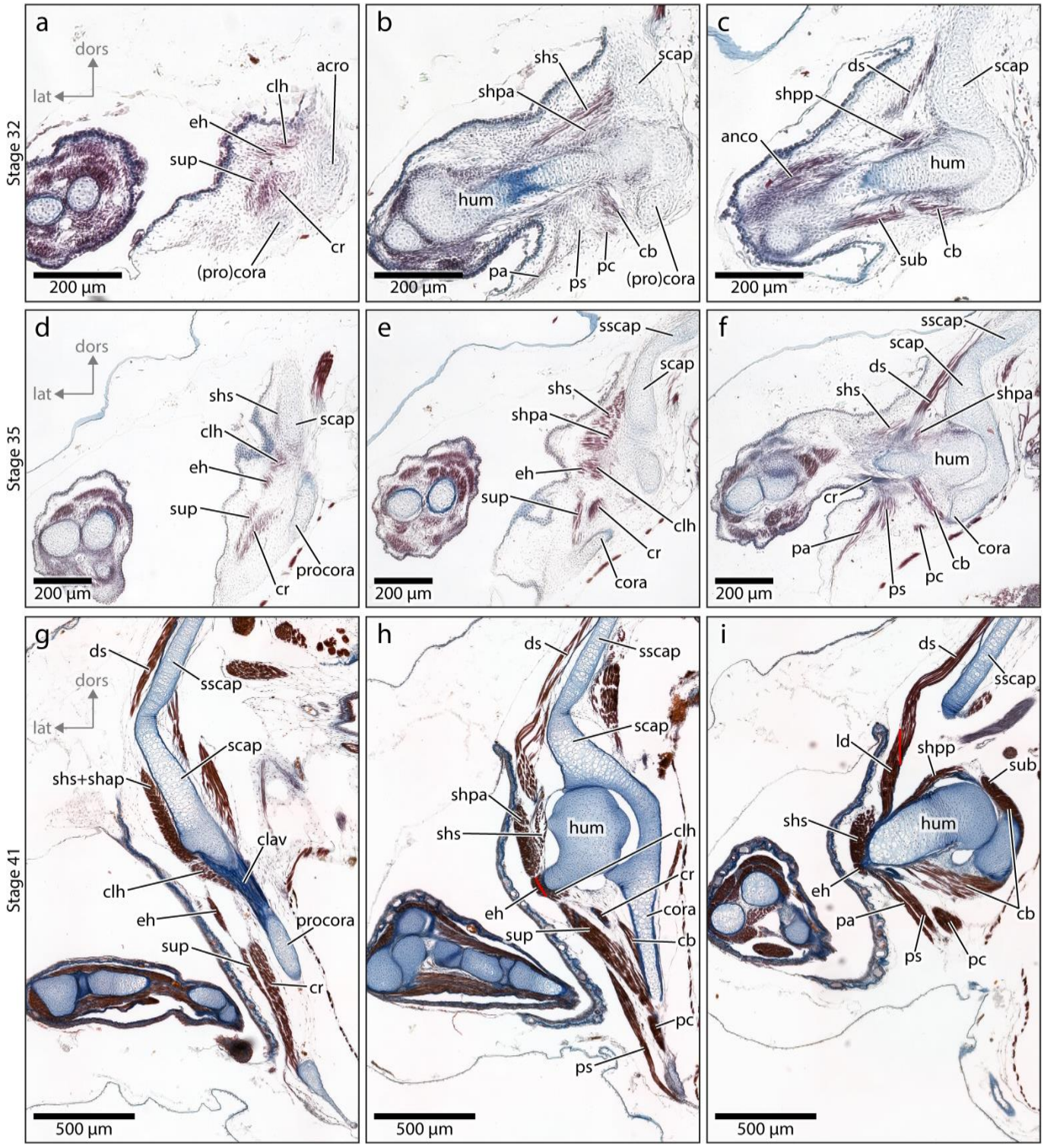


Fig. 8 Azan-stained histological sections of the shoulder joint region of *Bombina orientalis* larvae in different developmental stages. Left section is located anterior to the middle section, which, in turn, is anterior to the right. **a–c** Sections of Stage 32 (ZMH A12427), left-side shoulder joint, but mirrored to the right for consistency. **d–f** Sections of Stage 35 (ZMH A12429), right-side shoulder joint. **g–i** Sections of Stage 41 (ZMH A12435), right-side shoulder joint. Red line: separation of adjacent muscles. anco: heads of m. anconaeus not crossing the shoulder joint; cb: m. coracobrachialis; clav: clavicula; clh: m. cleidohumeralis; cora: coracoid; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; hum: humerus; ld: m. latissimus dorsi; lat: lateral; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; scap: scapula; sub: m. subcoracoscapularis; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sup: m. supracoracoideus; sscap: suprascapula.

(Fig. 8g-i). The m. supracoracoideus formed one muscle mass with no tendency towards the formation of two portions.

Development of the shoulder joint muscles in *Rana temporaria*

The developmental pattern of the shoulder joint muscles in *Rana temporaria* was similar to the patterns observed in the other two species, but the condensations of pre-muscle cells were less differentiated in the earliest developmental stage considered herein (Stage 32-33, ZMH A14736). Dorsal to the humerus, the precursor of the m. anconaeus was present as a condensation of pre-muscle cells, but the different heads of this muscle were inseparable. The mm. scapulohumeralis superficialis, profundus anterior and profundus posterior, dorsalis scapulae, and latissimus dorsi were present as a continuous pre-muscle mass (Fig. 9a-c); only the future common tendon of the mm. dorsalis scapulae and latissimus dorsi was recognizable as a region of comparably densely packed cells (Fig. 9c). Ventral to the humerus, there were two pre-muscle masses that were mostly continuous and only separated by the already recognizable posterior part of the future m. coracoradialis (Fig. 9b); anteriorly, the m. coracoradialis formed one pre-muscle mass with the future mm. cleidohumeralis, episternohumeralis, and supracoracoideus (Fig. 9a). The pre-muscle mass posterior to the m. coracoradialis represented the future mm. pectoralis, cutaneus pectoris, coracobrachialis, and subcoracoscapularis (Fig. 9b, c).

In the *R. temporaria* specimen of Stage 34 (ZMH A14739), the pre-muscle masses were somewhat more differentiated. Dorsally, the caput scapulare of the m. anconaeus was distinct at the origin and distally continuous with other heads of m. anconaeus. The mm. dorsalis scapulae and latissimus dorsi formed one pre-muscle mass that was clearly distinct from the group of scapulohumeralis muscles. The future mm. scapulohumeralis superficialis and profundus anterior were continuous (Fig. 9d, e), but separated from the m. scapulohumeralis posterior. The anterior of the two ventral pre-muscle masses observed in Stage 32-33 (ZMH A14736) was clearly separable from the posterior one and the m. coracoradialis became more distinct, but was still largely continuous with the common precursor of the future mm. cleidohumeralis, episternohumeralis, and supracoracoideus (Fig. 9e). Concerning the posteroventral pre-muscle mass, the portio abdominalis of the m. pectoralis was separated from the remaining pre-muscle mass at its origin (Fig. 9f). An anteriorly directed expansion of the pre-muscle mass represented the future m. cutaneus pectoris (Fig. 9e); the cell condensation preceding the m. cutaneus pectoris was posteriorly continuous with the future portionis abdominalis and sternalis of the m. pectoralis. The remaining continuous

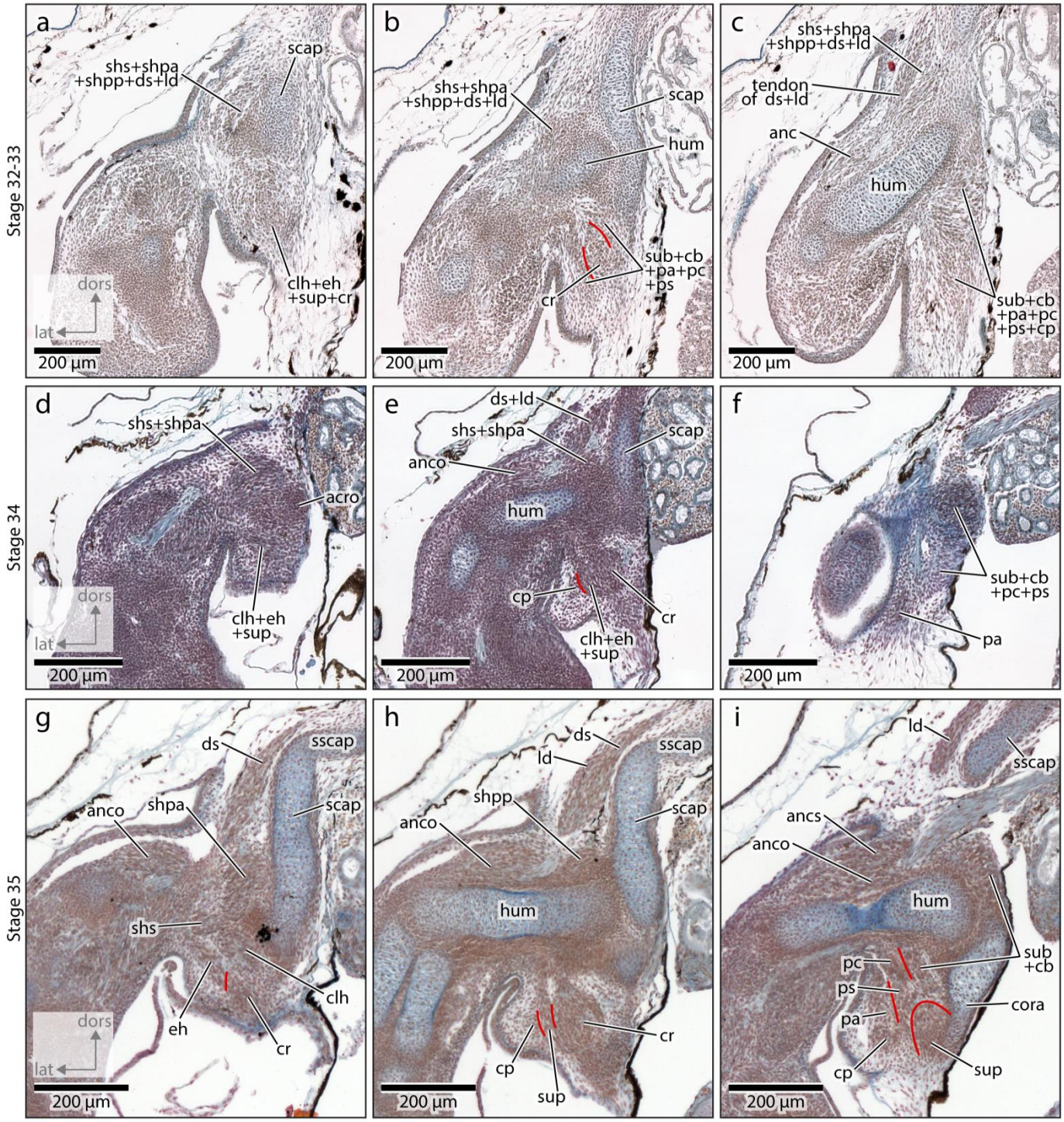


Fig. 9 Azan-stained histological sections of the right shoulder joint region of *Rana temporaria* larvae in different developmental stages. Left section is located anterior to the middle section, which, in turn, is anterior to the right. **a–c** Sections of Stage 32-33(ZMH A14736). **d–e** Sections of Stage 34 (ZMH A14739). **g–i** Sections of Stage 35 (ZMH A14740). Red line: separation of adjacent muscles. anc: m. anconaeus; anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; cb: m. coracobrachialis; clh: m. cleidohumeralis; cora: coracoid; cp: m. cutaneus pectoris; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; hum: humerus; ld: m. latissimus dorsi; lat: lateral; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; scap: scapula; sub: m. subcoracoscapularis; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sup: m. supracoracoideus; sscap: suprascapula.

posteroventral pre-muscle mass represented the portionis sternalis and coracoidea of the m. pectoralis and the mm. coracobrachialis and subcoracoscapularis.

In Stage 35 (ZMH A14740), the future mm. scapulohumeralis superficialis and profundus anterior were mostly continuous but formed two distinct heads at their insertions (Fig. 9g) and the m. latissimus dorsi was separated from the m. dorsalis scapulae. Anteroventrally, the mm. episternohumeralis and cleidohumeralis were separated from one (Fig. 9g) another and from the m. supracoracoideus; the latter muscle was more distinct from the m. coracoradialis (Fig. 9h). Posteroventrally, the portio abdominalis of m. pectoralis was entirely distinct from the other portions of this muscle, while the portionis sternalis and coracoidea were separated at their origins but remained continuous at their insertion (Fig. 9i). The future m. cutaneus pectoris was expanded anteriorly, but, posteriorly, remained continuous with the portionis abdominalis and sternalis of the m. pectoralis (Fig. 9h, i). The mm. coracobrachialis and subcoracoscapularis formed one continuous pre-muscle mass.

In the specimen of Stage 41 (ZMH A12870), the origin of the m. scapulohumeralis superficialis was shifted to the medial surface of the scapula and the anterodorsal part of the lateral coracoid. Thereby, the mm. scapulohumeralis superficialis and profundus anterior were separated at their origins (Fig. 10a, b). The m. cleidohumeralis was divided into two parts that were separated at their origins but continuous at their insertion; the pars profunda originated from the lateral part of the clavicula and the pars superficialis from the acromion (most anteroventral part of the scapula and the ventrally adjacent cartilage; Fig. 10a, b). The pars superficialis of the m. cleidohumeralis was closely associated with the mm. scapulohumeralis superficialis and profundus anterior at its origin (Fig. 10a). Posteriorly, all portions of the m. pectoralis were separated from one another and from the m. cutaneus pectoris (Fig. 10d). The m. coracobrachialis was split into a ventral and a dorsal portion (Fig. 10c) and the portio dorsalis of this muscle was closely associated and mostly continuous with the m. subcoracoscapularis (Fig. 10c, d).

Discussion

Muscle nomenclature and comparison to literature accounts

The muscles of the forelimbs in tetrapods are ontogenetically derived from two (ventral and dorsal) pre-muscle masses that form within the developing limb bud (summarized in Hirasawa and Kuratani 2018). These two pre-muscle masses split into the individual limb muscles during morphogenesis (Hirasawa and Kuratani 2018). Although none of the

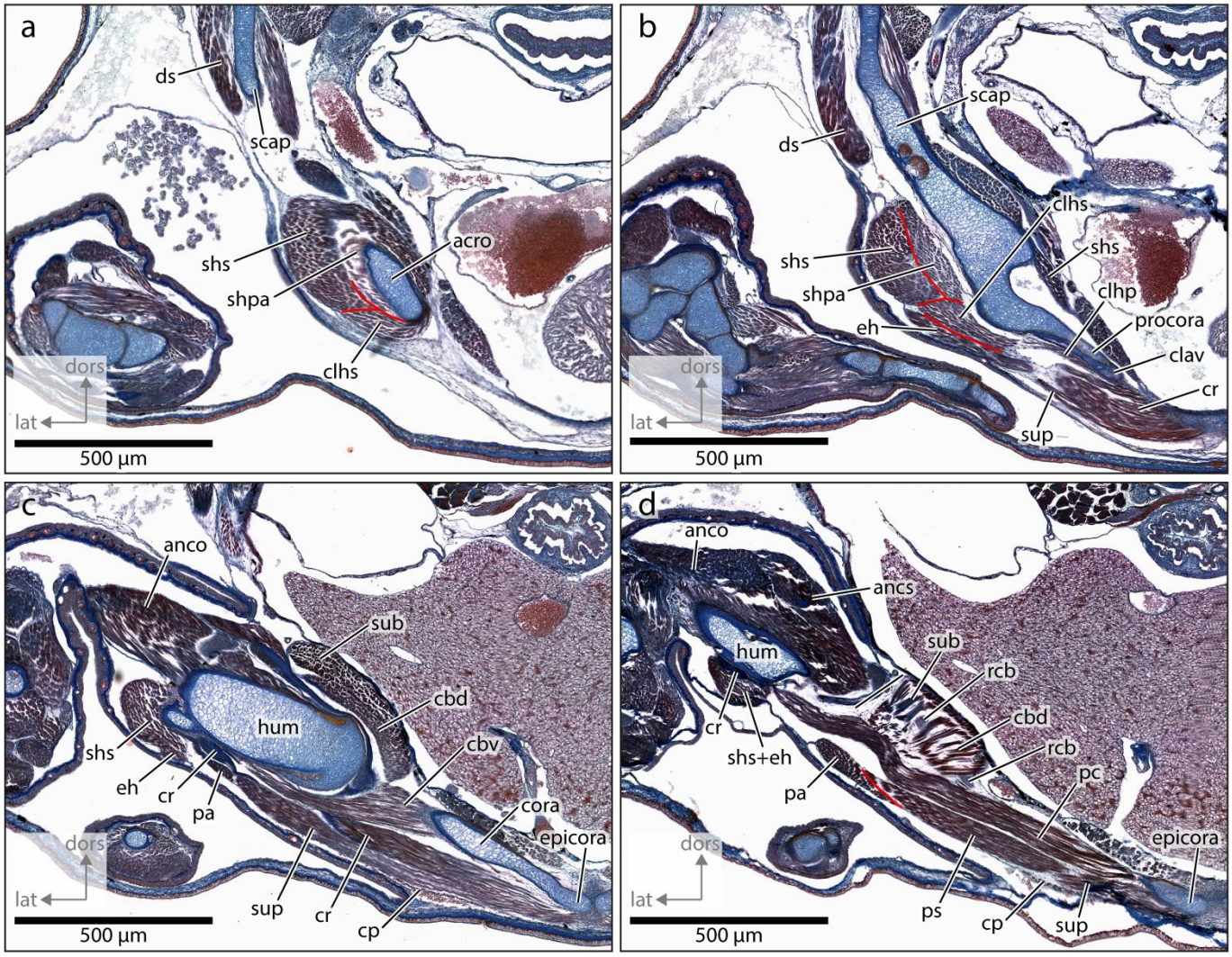


Fig. 10 Azan-stained histological sections of the right shoulder joint region of a Stage 41 larva of *Rana temporaria*. **a–d** Sections from anterior to posterior. Red line: separation of adjacent muscles. acro: acromion; anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; cbd: m. coracobrachialis pars dorsalis; cbv: m. coracobrachialis pars ventralis; clav: clavicula; clhp: m. cleidohumeralis pars profunda; clhs: m. cleidohumeralis pars superficialis; cora: coracoid; cp: m. cutaneus pectoris; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; pa: m. pectoralis portio abdominalis; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; rcb: r. coraco-brachialis; scap: scapula; sub: m. subcoracoscapularis; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; sup: m. supracoracoideus.

specimens examined herein showed only two undifferentiated pre-muscle masses, the observations presented herein are in accordance with the general pattern of limb muscle development in tetrapods: The various shoulder joint muscle entities present in late developmental stages or adult specimens ontogenetically originated from condensations of pre-muscle cells and formed by subdivisions of the latter. Extrapolating the observed pattern to earlier developmental stages, it seems likely that the mm. anconaeus, dorsalis scapulae,

latissimus dorsi, and the group of scapulohumeralis muscles are derived from a single dorsal pre-muscle mass hypothesized to be present in all tetrapods. Likewise, a single ventral pre-muscle mass presumably gives rise to the mm. cleidohumeralis, episternohumeralis, supracoracoideus, coracoradialis, subcoracoscapularis, coracobrachialis, and pectoralis (Fig. 11).

Regarding the muscles derived from the dorsal pre-muscle mass, our observations on the mm. anconaeus (all heads), dorsalis scapulae, and latissimus dorsi are in line with previous anatomical descriptions of *Ascaphus truei*, *Rana*, and *Bufo bufo* (Ritland 1955; Gaupp 1896; Bigalke 1927, respectively; Table 2). The m. scapulohumeralis superficialis usually is considered to be a part (pars scapularis sensu Bigalke 1927) of the m. deltoideus. The mm. episternohumeralis (pars cleido-humeralis longus sensu Bigalke 1927; pars episternalis sensu Gaupp 1896) and cleidohumeralis (pars clavicularis sensu Bigalke 1927) represent the other parts of the m. deltoideus in other studies; our results, however, indicate that the different muscle units that constitute the m. deltoideus in other studies, in fact, have different ontogenetic origins, namely either the ventral or the dorsal pre-muscle mass. Because of these different ontogenetic origins of the different parts, denoting them as parts of one muscle (m. deltoideus) might be misleading. Given that the muscle mass called m. scapulohumeralis superficialis herein is closely associated (ontogenetic development, innervation, and, in most species, continuity at origin) with the m. scapulohumeralis profundus anterior, it seems expedient and justified to discard its old name ‘m. deltoideus pars scapularis’. This also applies to the other muscles commonly considered to be a part of the m. deltoideus (suggested names summarized in Table 2).

Gaupp (1896) neither observed the m. scapulohumeralis profundus anterior, nor the m. scapulohumeralis profundus posterior in different species of *Rana*. Our analyses, however, confirmed the presence of both muscles in *R. temporaria* and support the notion of Bigalke (1927) that the m. deltoideus pars scapularis described by Gaupp (1896) also comprised the m. scapulohumeralis profundus anterior. Further support for this hypothesis can be found in Gaupp’s (1896) descriptions: He reported two separate origins and insertions for the pars scapularis and those origins correspond to the origins and insertions of the mm. scapulohumeralis superficialis and profundus anterior observed herein. In addition, Gaupp (1896) observed that some fibers of his m. deltoideus pars scapularis were innervated by the r. coraco-clavicularis, which contradicts our observations that this ramus only supplies muscles derived from the ventral pre-muscle mass. It seems likely that those fibers of Gaupp’s (1896) pars scapularis that are supplied by the r. coraco-clavicularis correspond to the pars

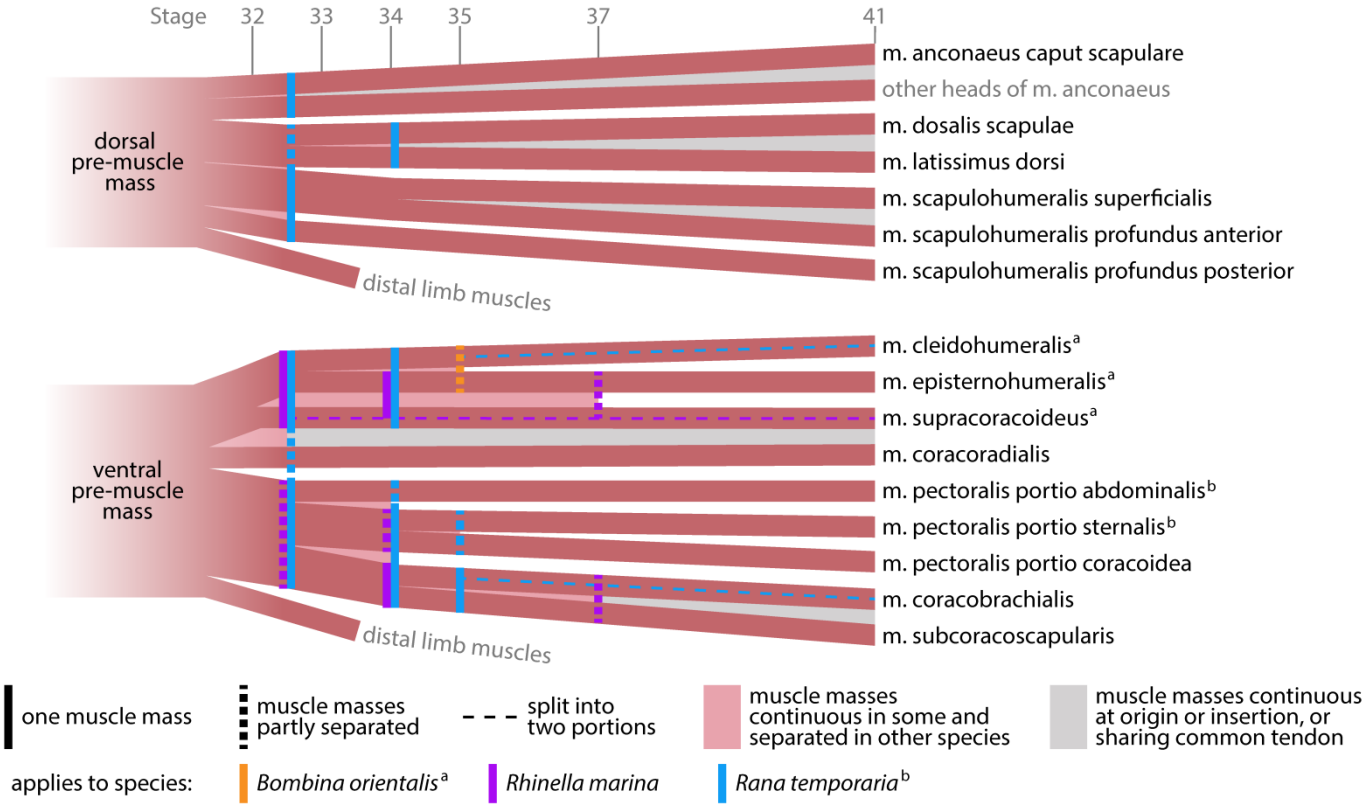


Fig. 11 Generalized ontogenetic splitting pattern of pre-muscle masses located ventral and dorsal to the humerus. Timelines derived from larvae of *Rhinella marina* (Stages 32-33, 34, 37, 41), *Rana temporaria* (Stages 32-33, 34, 35, 41), and *Bombina orientalis* (Stage 32 only). [a] *B. orientalis* only used to derive the splitting pattern of mm. cleidohumeralis, episternohumeralis, and supracoracoideus, as all other muscles were already distinct in the earliest stage examined, which is neglected in this figure. [b] In *R. temporaria*, the m. cutaneus pectoris separated from the portionis sternalis (not included in illustration) and abdominalis of the m. pectoralis.

superficialis of the m. cleidohumeralis due to the pattern of innervation and the observed close association of the pars superficialis with the m. scapulohumeralis superficialis in *R. temporaria*. If this was true, the pars scapularis of the m. deltoideus in Gaupp (1896) would comprise the mm. scapulohumeralis superficialis, scapulohumeralis profundus anterior, and the pars superficialis of the cleidohumeralis recognized herein. The pars clavicularis in Gaupp (1896) would be homologous to the pars profunda of the m. cleidohumeralis herein (Table 2), which would be consistent with the position and innervation of these muscles.

The single, undivided muscle mass (called m. deltoideus by Ritland 1955) observed in *Ascaphus truei* is formed by the mm. episternohumeralis, cleidohumeralis, and scapulohumeralis superficialis. The contribution of these latter three muscles to the formation of the former one muscle mass is supported by its absolute and relative position and by the muscle mass being supplied by the nerve branches that usually innervate these three muscles (rr. coraco-clavicularis and dorsalis scapulae anterior; own observations; Ritland 1955). Given

that the m. scapulohumeralis superficialis is ontogenetically derived from the dorsal, whereas the mm. episternohumeralis and cleidohumeralis are derived from the ventral pre-muscle mass, the most parsimonious explanation for the condition observed in *A. truei* is that the mm. episternohumeralis and cleidohumeralis did not separate during ontogenesis and are secondarily fused to the m. scapulohumeralis superficialis. This is supported by a previous observation that the corresponding muscle mass in *Leiopelma*, phylogenetically a close relative of *Ascaphus*, consisted of somewhat more distinct portions (Ritland 1955),

The muscle superficial (ventral) to the m. coracoradialis has previously been denoted as the pars epicoracoidea of the m. pectoralis (Gaupp 1896; Bigalke 1927) and as the m. supracoracoideus (Ritland 1955). We observed that the pre-muscle mass (anteroventral) that gave rise to this muscle separated from the pre-muscle mass (posteroventral) that gave rise to the other portions of the m. pectoralis early in ontogeny (Fig. 11). We suggest the use of the term ‘m. supracoracoideus’ instead of denoting this muscle as a portion of the m. pectoralis to highlight the split from the pectoralis group early in ontogeny (Table 2).

We further observed that the m. supracoracoideus split into an anterior and a posterior portion during the ontogenesis in *Rhinella marina*; there was no connection or association of this muscle to the m. coracobrachialis. The posterior portion of the m. supracoracoideus resembles the pars superficialis of the m. coraco-brachialis brevis in the descriptions illustrations of *Bufo bufo* by Bigalke (1927) with regard to the locations of the origin and the insertion on the skeletal elements and relative to other muscles . Therefore, we hypothesize that they are homologous. If so, the m. coracobrachialis would be present with only one portion in *R. marina* and *B. bufo*. In contrast, the ontogenetic pattern observed in *Rana temporaria* indicated that the m. coracobrachialis is present with two parts in that species: The separation of the two parts (dorsal and ventral) of the m. coracobrachialis occurred in a later developmental stage than the separation of their common precursor from the precursor of the m. supracoracoideus (Fig. 11). In the light of the evidence we consider it most parsimonious to assume that the partis dorsalis and ventralis, thus, truly are derivatives of the m. coracobrachialis in *R. temporaria* and neither of them is homologous to the muscle denoted m. coraco-brachialis brevis pars superficialis by Bigalke (1927) in *B. bufo*.

Ritland (1955) reported the presence of a superficial and a profound supracoracoideus muscle in *A. truei*. Our observations revealed that the muscle unit called m. supracoracoideus profundus by Ritland (1955) was located dorsal (profound) to the anastomosis formed by branches of the rr. coraco-clavicularis and coraco-brachialis. Given that a similar anastomosis was observed in *Alytes obstetricans* (Fig. 2) and *Bombina orientalis*, and that only the m.

coracobrachialis was located dorsal to the nerve branch in those two species, we suggest that the m. supracoracoideus profundus observed by Ritland (1955) in fact represents a part (pars ventralis) of the m. coracobrachialis. This interpretation is also consistent with the position of the insertion of mm. supracoracoideus and coracobrachialis relative to the m. coracoradialis: The fibers of the m. supracoracoideus were located ventral (superficial) to the m. coracoradialis and inserted onto the connective tissue covering the tendon of the latter whereas the fibers of the m. coracobrachialis lay dorsal (profound) to the m. coracoradialis and inserted onto the posterior surface of crista ventralis and the humerus (observed in *A. obstetricnas*, *B. orientalis*, *Rhinella marina*, and *Rana temporaria*; Figs 2, 6, Supporting Information Fig. S2). The insertions of the mm. supracoracoideus and coracobrachialis are, thus, separated by (the tendon of) the m. coracoradialis. If the muscle called m. supracoracoideus profundus by Ritland (1955) would be considered as a part of the m. coracobrachialis, the positions of the insertions of the mm. coracobrachialis and supracoracoideus relative to the m. coracoradialis would be identical to the conditions observed in the other species. If not, the insertion of the m. supracoracoideus profundus (sensu Ritland 1955) would have shifted compared to the other species.

We observed that the muscle unit, that has commonly been denoted as m. coraco-brachialis longus (Gaupp 1896; Bigalke 1927), was ontogenetically derived from the muscle precursor that also gave rise to the portionis sternalis and abdominalis of the m. pectoralis in *Rhinella marina* and likely also in *Rana temporaria*. We therefore consider this muscle to be a portion of the m. pectorals and suggest reflecting this relation by denoting it as the portio coracoidea of the m. pectoralis (Table 2). There were, however, some observations that render the suggested assignment to the m. pectoralis ambiguous: In some species (*Alytes obstetricans*, *Rana temporaria*) the portio coracoidea received nerve supply from the r. coraco-brachialis, and in *Ascalphusi truei* form the rr. coraco-brachialis and pectoralis proprius, which could imply a close association of the portio coracoidea of the m. pectoralis with the m. coraco-brachialis. Some ambiguity remains that requires more research but at the moment we consider the relationship to the m. pectoralis more plausible because of layer, position and ontogenetic origin.

The m. cutaneous pectoris was only present in *R. temporaria* and was ontogenetically derived from a shared anlage with the portionis sternalis and abdominalis of the m. pectoralis. This raises the question, if these portions of the pectoralis are homologous across the considered species or if such a homologization should include the m. cutaneous pectoris as some kind of de novo portion of the m. pectoralis.

Ritland (1955) described a m. subcoracoscapularis in *A. truei*. Neither Gaupp (1896), nor Bigalke (1927) described a muscle of this name in *Rana* and *Bufo bufo*, respectively. Gaupp (1896), however, observed that a similar muscle, that he denoted the pars profunda of m. corcaco-brachialis, was pierced by the r. coraco-brachialis and Bigalke (1927) reported two insertions (humeral spina tuberculi medialis and crista ventralis humeri) for the same muscle. Our results revealed the presence of the m. subcoracoscapularis in these species. Given that we observed the mm. subcoracoscapularis and coracobrachialis to always be continuous at their origins, that these two muscles were mostly continuous in *R. temporaria*, and that the r. coraco-brachialis passed between these two muscles, we assume that the pars profunda of m. coraco-brachialis described by Gaupp (1896) and Bigalke (1927) comprises the m. subcoracoscapularis (Table 2). If so, the pars profunda of m. coraco-brachialis described by Gaupp (1896) and Bigalke (1927) would not be homologous to the m. coracobrachialis herein. In order to avoid confusion we suggest referring to the parts of the m. coracobrachialis as pars dorsalis and pars ventralis, respectively.

Limitations

Our sampling of developmental stages was sparse and each of the developmental stages was represented by only one specimen per species. The observed differences in the timing of the developmental events (splitting of muscle units) may be subject to individual (within-species) variations and might not represent inter-specific variation. Yet, interspecific differences in the timing of the development of the pectoral girdle skeleton (Baleeva 2001) and its muscles (Soliz et al. 2018) have previously been reported for other anuran species.

The innervation patterns of the shoulder joint muscles were described using one specimen per species only and the observed differences between species might be caused by individual variations rather than species-specific peculiarities. The presence of sexual dimorphism has been observed for the humerus (Lee 2001; Padhye et al. 2015; Petrović et al. 2017) and certain muscle attached to the pectoral girdle (Oka et al. 1984; Emerson 1990; Lee 2001) in some anuran species. Sexual dimorphism and how the observed patterns are modified in a gender-specific fashion need further investigation.

Summary and conclusion

The anuran mm. anconaeus, dorsalis-scapulae, latissimus dorsi, coracoradialis, and the portionis sternalis and abdominalis of the m. pectoralis have consistently been recognized and denoted in the previous studies (Gaupp 1896; Bigalke 1927; Ritland 1955) and are reassessed

herein. The muscle unit called coraco-brachialis longus in previous studies has also been consistently recognized, but the name for this muscle is misleading as it suggests a close relation to the m. coracobrachialis (as used herein), whereas it is ontogenetically closely associated with the portions sternalis and abdominalis of the m. pectoralis. The name 'pectoralis portio coracoidea' seems more appropriate. The muscle entities that were previously considered as parts of a m. deltoideus ontogenetically arise from different pre-muscle masses (ventral and dorsal) observed in early development of the tetrapod limb bud (Hirasawa and Kuratani 2018). This composite nature of the 'm. deltoideus', in our opinion, warrants alternative terms (i.e., mm. scapulohumeralis superficialis, cleidohumeralis, episternohumeralis) that better clarify the independent nature of the muscle entities. The m. scapulohumeralis superficialis is closely associated with the m. supracoracoideus profundus anterior and both these muscles have occasionally been described as one muscle; the m. supracoracoideus profundus posterior likely has been overlooked by some authors. The m. subcoracoscaplualris, although mostly inseparable from the m. coracobrachialis in *Rana temporaria*, is present in all species examined and is characterized by being superficial to the m. coracobrachialis and the r. coraco-brachialis. The mm. cleidohumeralis, supracoracoideus, and coracobrachialis are present with two parts or portions in some species, these portions have not always been recognized and assigned correctly in previous studies.

In our study we applied homology criteria to sort and clarify the inconsistencies in the literature. Bearing the same term is not a prerequisite *per se* to establish homology statement for two entities. It is our experience, however, that the previously applied terms and contradictions in the literature obstructed the understanding of the evolution of the shoulder joint muscles. The terminology we propose mostly recruits from existing terms and tries to limit changes to the necessary. It clarifies, in our opinion, many discrepancies, and offers more parsimonious explanations of the observed patterns than previous systems. This case study highlights the importance of critically questioning published anatomical descriptions before they were to be used for comparisons in other studies. For example, the observation that the muscles commonly considered to be parts of the m. deltoideus are ontogenetically derived from different pre-muscle masses renders the overall homologization of 'the m. deltoideus' with single muscles or muscle complexes in other vertebrate taxa questionable. In our opinion, only well-supported homologization and primary homology statements of muscle units across taxa allows for the reliable reconstruction of ancestral character states and should be scrutinized before they were used to derive evolutionary hypotheses or conclusions.

Acknowledgements

We are grateful to Katharina Ruthsatz for providing larvae of Rana temporaria, André Beerlink and YXLON International GmbH for performing the µCT scans of Bufo bufo, and Thomas Kleinteich and the TPW Prüfzentrum GmbH for performing the µCT scan of Rhinella marina. In addition, we thank Frank Friedrich for fruitful discussions, and Carla Viola Reinbold and Dilara Leptin for their help with digitizing the histological serial sections. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 387723284.

Authors' contributions

KE conceived the study, curated and analyzed the data, performed the 3d reconstructions, and drafted the manuscript. KE and SP generated the volume data. KE, SP, and AH discussed the results. AH provided resources, edited an early version of the manuscript, and acquired funding. All authors critically revised the manuscript and approved the final version.

Declarations

Funding

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 387723284.

Conflicts of interest

The authors declare that they have no conflict of interest.

Ethics approval

This study does not contain any experiments with human participants or living animals performed by any of the authors.

Consent to participate

Not applicable.

Data availability

The volume data generated for the study will be uploaded to https://www.fdr.uni-hamburg.de.

Code availability

Not applicable.

References

Abdala V, Diogo R (2010) Comparative anatomy, homologies and evolution of the pectoral and forelimb musculature of tetrapods with special attention to extant limbed amphibians and reptiles. J Anat 217:536–573. https://doi.org/10.1111/j.1469-7580.2010.01278.x

Baleeva NV (2001) Formation of the scapular part of the pectoral girdle in anuran larvae. Russ J Herpetol 8:195–204

Baleeva NV (2009) Formation of the coracoid region of the anuran pectoral girdle. Russ J Herpetol 16:41–50

Bigalke R (1927) Zur Myologie der Erdkröte. (*Bufo vulgaris*, Laurenti.). Z Anat Entwickl Gesch 82:286–353. https://doi.org/10.1007/BF02119515

Burton TC (1983) The musculature of the papuan frog *Phrynomantis stictogaster* (Anura, Microhylidae). J Morphol 175:307–324. https://doi.org/10.1002/jmor.1051750308

Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M, Tomancak P, Hartenstein V, Douglas RJ, Samuel A (2012) TrakEM2 software for neural circuit reconstruction. PLoS ONE 7:e38011. https://doi.org/10.1371/journal.pone.0038011

Coates MI, Jeffery JE, Rut M (2002) Fins to limbs: what the fossils say. Evol Dev 4:390–401. https://doi.org/10.1046/j.1525-142X.2002.02026.x

Cunningham DJ (1890) The value of nerve-supply in the determination of muscular homologies and anomalies. J Anat Physiol 25:31–40

de Pinna MCC (1991) Concepts and tests of homology in the cladistic paradigm. Cladistics 7:367–394. https://doi.org/10.1111/j.1096-0031.1991.tb00045.x

Diogo R, Abdala V (2007) Comparative anatomy, homologies and evolution of the pectoral muscles of bony fish and tetrapods: a new insight. J Morphol 268:504–517. https://doi.org/10.1002/jmor.10531

Diogo R, Ziermann JM (2014) Development of fore- and hindlimb muscles in frogs: morphogenesis, homeotic transformations, digit reduction, and the forelimb-hindlimb enigma. J Exp Zool B Mol Dev Evol 322:86–105. https://doi.org/10.1002/jez.b.22549

Diogo R, Johnston P, Molnar JL, Esteve-Altava B (2016) Characteristic tetrapod musculoskeletal limb phenotype emerged more than 400 MYA in basal lobe-finned fishes. Sci Rep 6:37592. https://doi.org/10.1038/srep37592

Duellman WE, Trueb L (1994) Biology of amphibians, Johns Hopkins Paperbacks edition. Johns Hopkins University Press, Baltimore

Emerson SB (1990) Scaling of the epicoracoid horn muscle in arciferal frogs. J Herpetol 24:84–87

Engelkes K, Friedrich F, Hammel JU, Haas A (2018) A simple setup for episcopic microtomy and a digital image processing workflow to acquire high-quality volume data and 3D surface models of small vertebrates. Zoomorphology 137:213–228. https://doi.org/10.1007/s00435-017-0386-3

Engelkes K, Kath L, Kleinteich T, Hammel JU, Beerlink A, Haas A (2020, accepted) Ecomorphology of the pectoral girdle in anurans (Amphibia, Anura): shape diversity and biomechanical considerations. Ecol Evol

Gaupp EWT (1896) A. Ecker's und R. Wiedersheim's Anatomie des Frosches. Erste Abteilung. Lehre vom Skelett und vom Muskelsystem, 3rd edn. Friedrich Vieweg und Sohn, Braunschweig

Gaupp EWT (1899) A. Ecker's und R. Wiedersheim's Anatomie des Frosches. Zweite Abteilung. Lehre vom Nerven- und Gefässsystem, 2nd edn. Friedrich Vieweg und Sohn, Braunschweig

Gegenbaur C (1865) Untersuchungen zur vergleichenden Anatomie der Wirbelthiere: Zweites Heft - Schultergürtel der Wirbelthiere, Brustflosse der Fische. Wilhem Engelmann, Leipzig

Gosner KL (1960) A simplified table for staging anuran embryos and larvae with notes on identification. Herpetologica 16:183–190

Haines RW (1935) A consideration of the constancy of muscular nerve supply. J Anat 70:33–55

Havelková P, Roček Z (2006) Transformation of the pectoral girdle in the evolutionary origin of frogs: insights from the primitive anuran *Discoglossus*. J Anat 209:1–11. https://doi.org/10.1111/j.1469-7580.2006.00583.x

Hirasawa T, Kuratani S (2018) Evolution of the muscular system in tetrapod limbs. Zoological Lett 4:2107. https://doi.org/10.1186/s40851-018-0110-2

Holliday CM, Witmer LM (2007) Archosaur adductor chamber evolution: integration of musculoskeletal and topological criteria in jaw muscle homology. J Morphol 268:457–484. https://doi.org/10.1002/jmor.10524

Jetz W, Pyron RA (2018) The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat Ecol Evol 2:850–858. https://doi.org/10.1038/s41559-018-0515-5

Jones EI (1933) Observations on the pectoral musculature of Amphibia Salientia. Ann Mag Nat Hist 12:403–420. https://doi.org/10.1080/00222933308673704

Kerr NS (1955) The homologies and nomenclature of the thigh muscles of the opossum, cat rabbit, and rhesus monkey. Anat Rec 121:481–493. https://doi.org/10.1002/ar.1091210302

Krell F-T, Cranston PS, Krell F-T (2004) Which side of the tree is more basal? Syst Entomol 29:279–281. https://doi.org/10.1111/j.0307-6970.2004.00262.x

Lee J (2001) Evolution of a secondary sexual dimorphism in the toad, *Bufo marinus*. Copeia 2001:928–935. https://doi.org/10.1643/0045-8511(2001)001[0928:EOASSD]2.0.CO;2

Maglia AM, Púgener LA (1998) Skeletal development and adult osteology of *Bombina orientalis* (Anura: Bombinatoridae). Herpetologica 54:344–363

Mahendra BC (1936) A case of polymely in the Indian bull-frog, *Rana tigrina* Daud. Proc Indian Natl Sci Acad B Biol Sci 4:483–493. https://doi.org/10.1007/BF03051421

Manzano A, Abdala V, Herrel A (2008) Morphology and function of the forelimb in arboreal frogs: specializations for grasping ability? J Anat 213:296–307. https://doi.org/10.1111/j.1469-7580.2008.00929.x

McGonnell IM (2001) The evolution of the pectoral girdle. J Anat 199:189–194. https://doi.org/10.1046/j.1469-7580.2001.19910189.x

Metscher BD (2009) MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. BMC Physiol 9:1–14. https://doi.org/10.1186/1472-6793-9-11

Minkoff EC (1974) The Fürbringer hypothesis of nerve–muscle specificity reexamined with respect to the facial musculature. Can J Zool 52:525–532. https://doi.org/10.1139/z74-065

Molnar JL, Diogo R, Hutchinson JR, Pierce SE (2018) Reconstructing pectoral appendicular muscle anatomy in fossil fish and tetrapods over the fins-to-limbs transition. Biol Rev 93:1077–1107. https://doi.org/10.1111/brv.12386

Oka Y, Ohtani R, Satou M, Ueda K (1984) Sexually dimorphic muscles in the forelimb of the Japanese toad, *Bufo japonicus*. J Morphol 180:297–308. https://doi.org/10.1002/jmor.1051800310

Omland KE, Cook LG, Crisp MD (2008) Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. Bioessays 30:854–867. https://doi.org/10.1002/bies.20794

Padhye AD, Jadhav A, Sulakhe S, Dahanukar N (2015) Sexual dimorphism in the Kudremukh Bush Frog (Anura: Rhacophoridae: *Raorchestes tuberohumerus*) of the Western Ghats, India, with a note on its distribution and conservation status. J Threat Taxa 7:7211–7222. https://doi.org/10.11609/JoTT.o4192.7211-22

Petrović TG, Vukov TD, Tomašević Kolarov N (2017) Morphometric ratio analyses: locomotor mode in anurans. C R Biol. https://doi.org/10.1016/j.crvi.2017.02.004

Ponomartsev S, Valasek P, Patel K, Malashichev Y (2017) Neural crest contribution to the avian shoulder girdle and implications to girdle evolution in vertebrates. Biol Commun 62:26–37. https://doi.org/10.21638/11701/spbu03.2017.104

Púgener LA, Maglia AM (1997) Osteology and skeletal development of *Discoglossus sardus* (Anura: Discoglossidae). J Morphol 233:267–286. https://doi.org/10.1002/(SICI)1097-4687(199709)233:3<267:AID-JMOR6>3.0.CO;2-0

Pyron AR, Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol 61:543–583. https://doi.org/10.1016/j.ympev.2011.06.012

Remane A (1952) Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Theorerische Morphologie und Systematik I. Akademische Verlagsgesellschaft Geest & Portig, Leipzig

Ritland RM (1955) Studies on the post-cranial morphology of *Ascaphus truei*. II. Myology. J Morphol 97:215–282. https://doi.org/10.1002/jmor.1050970203

Rolleston G (1869) On the homologies of certain muscles connected with the shoulder-joint. Trans Linn Soc Lond 26:609–629. https://doi.org/10.1111/j.1096-3642.1869.tb00541.x

Romer AS (1924) Pectoral limb musculature and shouldergirdler structure in fish and tetrapods. Anat Rec 27:119–143. https://doi.org/10.1002/ar.1090270210

Romer AS (1922) The locomotor apparatus of certain primitive and mammal-like reptiles. Bull Am Mus Nat Hist 46:517–606

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676–682. https://doi.org/10.1038/nmeth.2019

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089

Shearman RM (2005) Growth of the pectoral girdle of the Leopard frog, *Rana pipiens* (Anura: Ranidae). J Morphol 264:94–104. https://doi.org/10.1002/jmor.10322

Shearman RM (2008) Chondrogenesis and ossification of the lissamphibian pectoral girdle. J Morphol 269:479–495. https://doi.org/10.1002/jmor.10597

Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. Nature 457:818–823. https://doi.org/10.1038/nature07891

Shubin NH, Daeschler EB, Jenkins FA (2006) The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. Nature 440:764–771. https://doi.org/10.1038/nature04637

Soliz MC, Ponssa ML, Abdala V (2018) Comparative anatomy and development of pectoral and pelvic girdles in hylid anurans. J Morphol 217:536. https://doi.org/10.1002/jmor.20820

Tyson H (1987) The structure and development of the anuran breast-shoulder apparatus, forelimb, and associated musculature. PhD Thesis, University of Alberta

Zachos FE (2016) Tree thinking and species delimitation: guidelines for taxonomy and phylogenetic terminology. Mamm Biol 81:185–188. https://doi.org/10.1016/j.mambio.2015.10.002

Zbären J (1966) Eine Azan-ähnliche Färbung mit Kernechtrubin als Kernfarbstoff. Mikrokosmos 55:286

Supporting Information

[The following Supporting Information can be found at the end of this thesis, pp. 281-283.]

Fig. S1 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**, **g**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**, **h**).

Fig. S2 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**).

Fig. S3 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**).

- Chapter seven -

General discussion

The results of Chapters two and three show that very accurate surfaces and landmark data can be derived from CT volumes of the pectoral girdle bones in anurans if suitable methodological approaches are used. This gives confidence that the observations and conclusions regarding the ecomorphology and function of the pectoral girdle are reliable (Chapter four). The phylogenetic relationships, size, and locomotor behavior have an effect on the shape of the pectoral girdle in anurans, but there are differences in the relative impact of these factors between the different bones. A remarkable diversity of pectoral girdle shapes has evolved within locomotor groups indicating that different shapes allow for similar functions (many-to-one mapping of form onto function). Hypothetical considerations on the evolution of the observed within-group shape diversity will be discussed below.

Significant shape differences have mainly been related to the overall pectoral girdle geometry and the shape of the coracoid. The most prominent shape differences have been observed between burrowing and non-burrowing species with headfirst and backward burrowing species significantly differing from one another and from the other locomotor groups. The pectoral girdle shapes of burrowing species have generally larger moment arms for (simulated) humerus retractor muscles across the shoulder joint, which might be an adaptation to and a biomechanical necessity for burrowing. The mechanisms of how the moment arms were enlarged differed between species and were associated with differences in the reaction of the coracoid to simulated loading by physiologically relevant forces. However,

the analyses on the efficiency of the humerus retractor muscles in Chapter four only took the effects of the pectoral girdle geometry into account; the impact of the humerus on the moment arms of retractor muscles was neglected. The latter will be considered in the following by discussing the properties of a lever-force-system modeling the anuran shoulder joint.

The comparison of the actual configurations of the shoulder joint muscles across anuran species and the revealing of potential adaptations to locomotor behaviors requires the knowledge of inter-specific muscle homologies. Therefore, inconsistencies in the identification, homologization, and naming of shoulder joint muscles found in the literature have been resolved in Chapter six. Below, the shoulder joint muscle morphology observed in anurans will be compared to the muscles in salamanders in order to derive hypothesis on the inter-order homology of muscle units and to provide a base for potential future studies that fit the Anura within the larger picture of higher-order homologies across vertebrates and muscle or forelimb evolution. Finally, the morphological methods used in this study and the role of illustrations in communicating anatomical observations will be discussed at the end of this chapter.

Evolution of the pectoral girdle shape in the context of locomotoion

The geometric morphometric analyses in Chapter four revealed a significant phylogenetic component in the shapes of the pectoral girdle bones of anurans. This means that distantly related species differ in their pectoral girdle shapes, whereas closely related species have similarly shaped pectoral girdles just because of their, respectively, distant or close relatedness. Considering this and that a combination of walking and hopping or jumping is the ancestral mode of locomotion for anurans (Přikryl et al. 2009; Reilly & Jorgensen, 2011), the remarkable shape diversity within the groups of walking/hopping and jumping species observed in Chapter four indicates that only a minor or even no selective pressure toward a specific shape was imposed by the biomechanical requirements of those locomotor behaviors. The different pectoral girdle shapes seem to be equally suitable for walking, hopping, or jumping (many-to-one mapping; Wainwright et al. 2005) and the shapes likely evolved mostly randomly within these two locomotor groups.

Climbing, backward burrowing, and headfirst burrowing behavior each evolved several times independently within the Anura (Nomura et al. 2009; Reilly & Jorgensen, 2011; Keeffe & Blackburn, 2020). The lineages, in which one or more of these derived locomotor modes evolved, likely differed with regard to the shapes of their pectoral girdle bones, because

phylogeny had a significant effect on the shape. Consequently, the potential selection for specific anatomical adaptations to evolutionary new locomotor behavior acted on disparate ancestral girdle geometries and, therefore, likely resulted in different adaptations to similar locomotor behaviors. The results reported in Chapter four do not support this hypothesis for adaptations to climbing, but they do so for burrowing. The pectoral girdle shapes of backward and headfirst borrowing species significantly differed from one another and from the shapes observed in other locomotor groups. In addition, the pectoral girdle shapes of burrowing species had specific biomechanical properties that are advantageous at least for headfirst burrowing, but these advantageous properties were facilitated by different mechanisms: The moment arms of the humerus retractor muscles were either enlarged by a posteriorly directed coracoid (e.g., *Hemisus marmoratus*) or by the presence of a sternum (e.g., *Sphaerotheca breviceps*) and the coracoids showed specie-specific patterns of von Mises stress implying different functions in the transmission and dissipation of forces. These different solutions for the same 'problem' of increasing the moment arms of the humerus retractor muscles might have evolved because selection acted on different ancestral girdle shapes.

Efficiency of humerus retraction in the context of burrowing

The analyses in Chapter four indicate that an efficient humerus retraction is advantageous for burrowing species. In that context, only the effect of the pectoral girdle geometry on the moment arms of a hypothetical humerus retractor muscle was considered. Previous conclusions derived from lever systems modelling the musculoskeletal system of the forelimb (Emerson, 1991) indicate that also the length of the humerus and the locations of the insertion of the retractor muscles along this length influence a specimen's performance in moving the humerus. This assumption is also supported by recent studies that reported the humerus shape of headfirst burrowing species to differ from the humerus shape in other species (Keeffe & Blackburn, 2020) and the shape of limb bones (fore- and hindlimb) to differ between most locomotor groups (Stepanova & Womack, 2020).

The humerus can be modeled as a Class 3 lever (as defined in Davidovits, 2008): The humerus (lever) is free to rotate about the shoulder joint (fulcrum) if forces produced by muscles or external loads act on it. Simplifying this model to represent only the case of humerus retraction allows for omitting the third dimension (ventral-dorsal axis) and for the reduction of the humerus retractor muscles to the hypothetical posterior muscle defined in Chapter four (Fig. 1A, B).

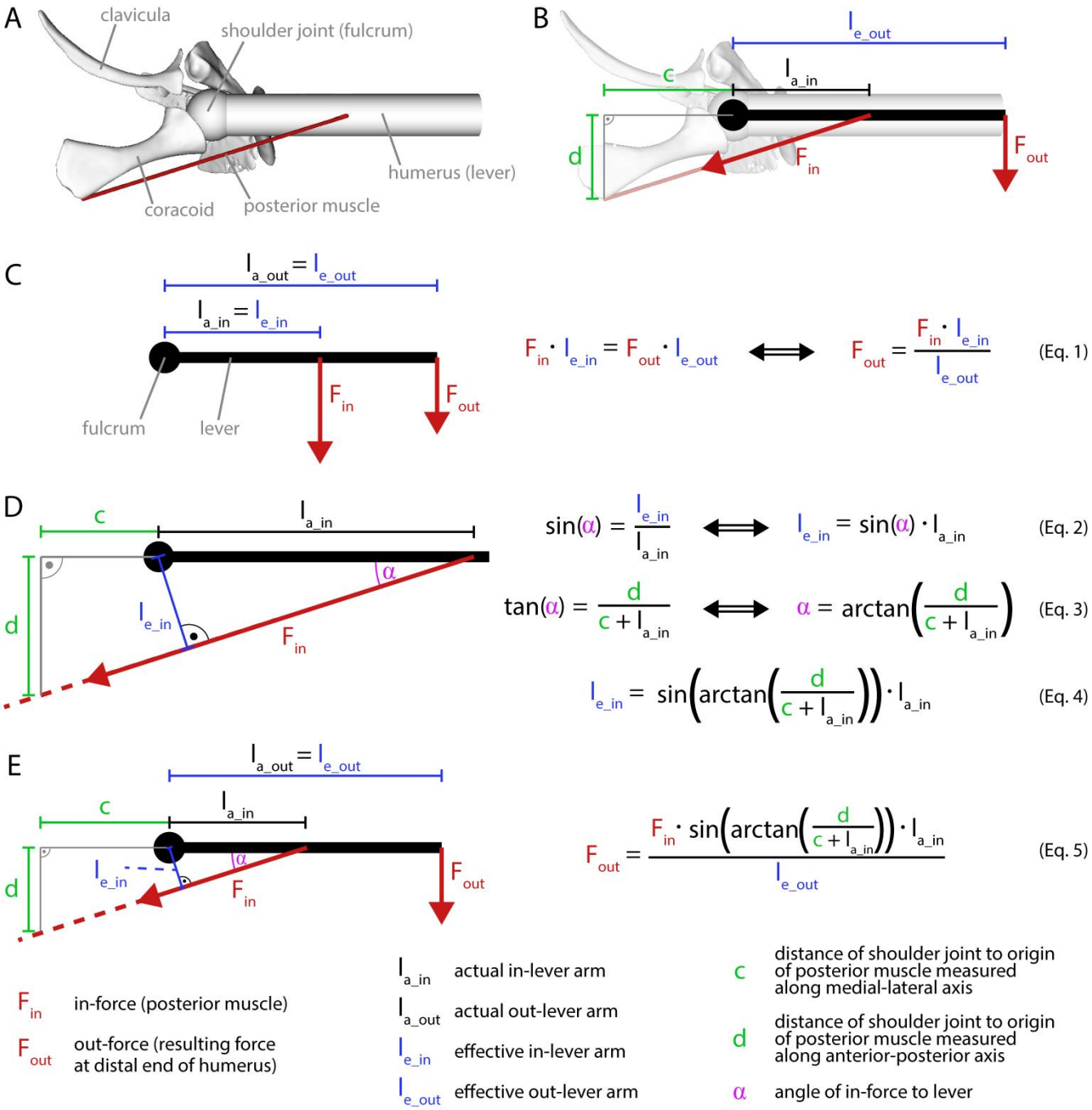


Figure 1 Class 3 lever and forces modelling humerus retraction in anurans. **A** Musculoskeletal model of *Ecnomiohyla miliaria* created in Chapter four reduced to posterior muscle. **B** Schematic illustration of lever and relevant variables that determine the resulting force at the distal end of the humerus. **C** Lever with in- and out-force acting perpendicular to the lever and equation describing the relation of forces and lever arms (Eq. 1). **D** Reduced model showing only an in-force acting obliquely on the lever and the associated actual and effective in-lever arms, as well as distances (c, d) describing the position of the insertion of the posterior muscle relative to the shoulder joint. Equations model the dependence of the effective in-lever arm on the angle of the in-force to the humerus (Eq. 2), the dependence of the angel on the position of the origin and insertion of the posterior muscle (Eq. 3), and the resulting dependence of the effective in-lever arm on the position of the origin and insertion of the posterior muscle relative to the shoulder joint (Eq. 4). **E** Full model of humerus retraction and equation for calculating the out-force given the in-force, the humerus length, and the position of the origin and insertion of the posterior muscle relative to the shoulder joint (Eq. 5).

If a force (in-force) acts perpendicularly on a lever and the resulting force (out-force) is also measured perpendicular to the lever, the out-force can immediately be calculated from the in-force and the distances of the points of force application to the fulcrum (i.e., the lever arms; Hildebrand, 1995; Fig. 1C: Eq. 1). If a force acts obliquely to the lever, the distance of its point of application (actual lever arm) does not equal the effective lever arm. The effective lever arm (also called moment arm; Davis, 1974) equals the distance of the fulcrum to the closest point on a line in the direction of the force passing through the point of force application (Hildebrand, 1995). The actual lever arm, the effective lever arm, and the force's line of action, thus, form a right-angled triangle, and the effective lever arm can be calculated from the actual lever arm and the angle of the force to the lever (sine function; Fig. 1D: Eq. 2).

In the case of humerus retraction, the angle of the force applied to the humerus equals the line of action of the posterior muscle. The muscle's line of action, in turn, depends on the locations of the origin and the insertion relative to the shoulder joint. This relation allows for the calculation of the angel between the force and the humerus (tangent function; Fig. 1D: Eq. 3). In total, the effective in-lever arm can be calculated if the locations of the origin and insertion of the posterior muscle relative to the shoulder joint are given (substitution of Eq. 3 into Eq. 2; Fig. 1D: Eq. 4). Consequently, the out-force generated at the distal end of the humerus (without loss of generality measured perpendicular to the humerus) by the contraction of the posterior muscle depends on the magnitude of the in-force, the length of the humerus, and the locations of the origin and insertion of the posterior muscle relative to the shoulder joint (substitution of Eq. 4 into Eq. 1; Fig. 1E: Eq. 5).

Separately testing the effect of each variable in the lever system on the magnitude of the out-force using Equation 5 (Fig. 1E; calculations performed in RStudio version 1.1.463 based on R version 3.5.3; R Core Team, 2019; RStudio Team, 2018) reveals that the resulting force at the distal end of the humerus can be increased by an increase in the force produced by the retractor muscle(s), a shortening of the humerus, a more lateral (distal) insertion of the retractor muscle(s) on the humerus, a shortening of the distance between the origin of the posterior muscle and the shoulder joint along the medial-lateral axis, and a posteriorly shifted origin of the posterior muscle relative to the shoulder joint (Fig. 2).

The realization of most of the mechanism to increase the out-force in burrowing species has been reported, or at least been indirectly observed in previous studies. Some headfirst burrowing species have enlarged humerus retractor muscles (Emerson, 1976; Keeffe & Blackburn, 2020). These enlarged muscles potentially produce higher in-forces acting on the

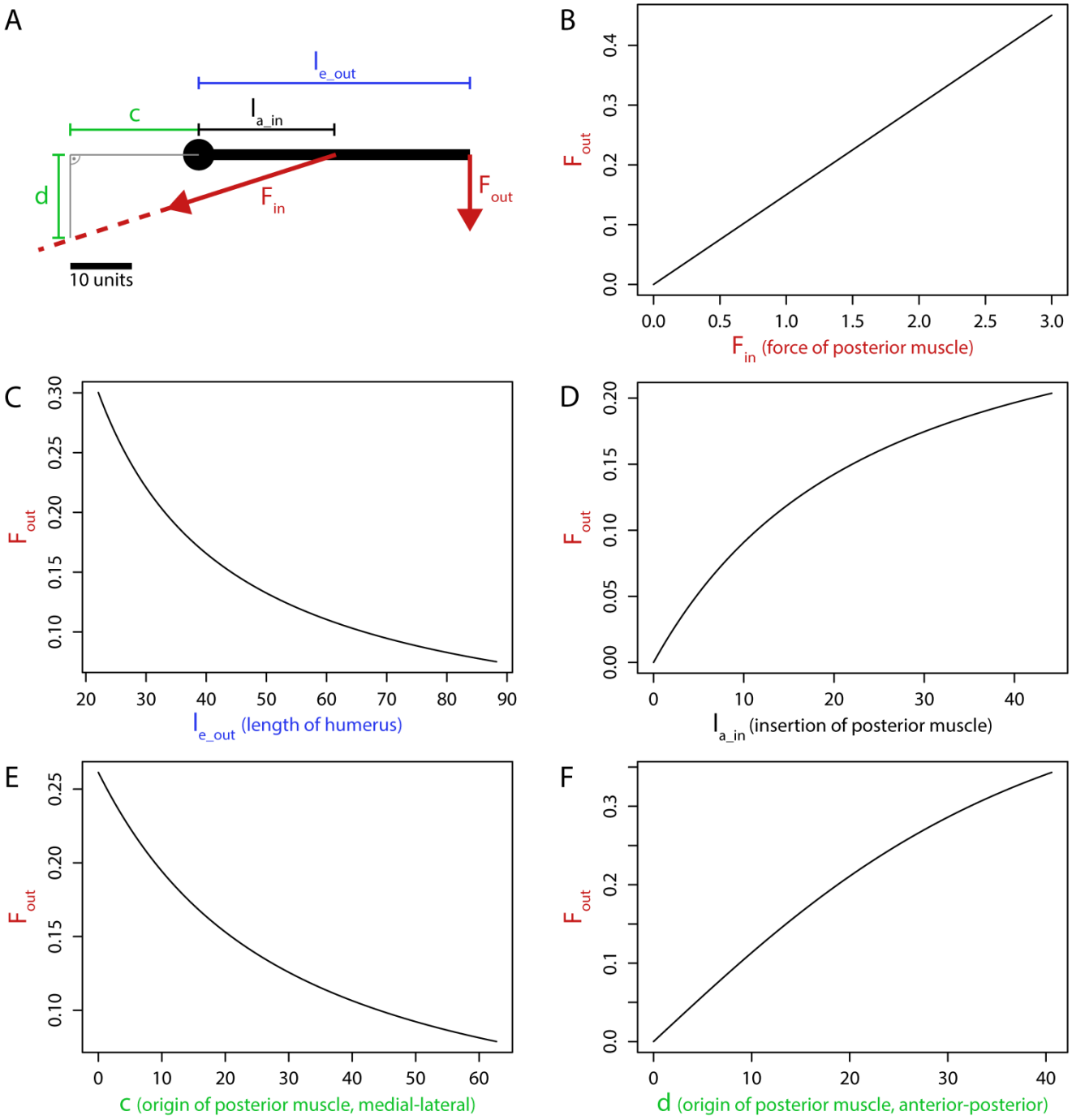


Figure 2 Effects of magnitude of in-force, humerus length, and locations of the origin and insertion of the posterior muscle relative to the shoulder joint on the out-force generated at the distal end of the humerus. Units arbitrary, but consistent across plots. **A** Leve, forces, and distances modelling the humerus and a hypothetical retractor muscle (same model as in Fig. 1B, E). **B** Effect of varying in-forces (produced by posterior muscle) on the out-force generated at the distal end of the humerus, whilst all other variables are kept constant. **C** Effect of varying effective out-lever arms (length of the humerus) on the out-force, other variables constant. **D** Effect of varying actual in-lever arms (location of the insertion of the posterior muscle relative to the shoulder joint) on the out-force, other variables constant. **E** Effect of varying (along medial-lateral axis) locations of the origin of the posterior muscle relative to the shoulder joint on the out-force, other variables constant. **F** Effect of varying (along anterior-posterior axis) locations of the origin of the posterior muscle relative to the shoulder joint on the out-force, other variables constant.

humerus, as the muscle mass is linked to the physiological cross-section area, which in turn is proportional to the maximal force a muscle can produce (Lieber & Fridn, 2000). In addition, the crista ventralis of the humerus, that is, the area for the insertion of the humerus retractor muscles, is comparably large in burrowing species if compared to non-burrowing species (Emerson, 1976; Helfsgott, 2020 [Master Thesis]). This might indicate that also backward burrowing species have comparably large retractor muscle. The lateral expansion of the crista ventralis humeri observed in burrowing species potentially also contributes to increasing the force produced at the distal end of the humerus (Helfsgott, 2020 [Master Thesis]): The insertion of some of the retractor muscles might have been shifted to a more lateral (distal) point, which in turn would increase the out-force. The long bones of the limbs of burrowing species were found to be short if compare to non-burrowing species (Stepanova & Womack, 2020). A short humerus corresponds to a short out-lever and thereby increases the out-force at the distal end of the humerus. Last but not least, the results in Chapter four revealed that the origin of the (hypothetical) humerus retractor muscle was shifted posteriorly in burrowing species, which resulted in a larger moment arm (effective in-lever arm) across the shoulder joint. The theoretical predictions derived from the lever model show that this results in an increase of the out-force produced at the distal end of the humerus.

The conclusions in this section are primary derived from theoretical considerations derived from a very reduced lever model of the shoulder joint in anurans. Although the realization of some of the mechanisms to increase the force produced at the distal end of the humerus has been (indirectly) observed in previous studies, the conclusions need to be further supported by future studies. Most importantly, the humerus retractor muscles in burrowing species need to be considered and compared to those in non-burrowing species as the muscle geometry was only hypothesized herein (see hypothetical posterior muscle).

Predictions for the biomechanics of humerus protraction and adduction in the context of landing

Various landing strategies ranging from ‘belly-flops’ to coordinated landing on the forelimbs have evolved within the Anura (Essner et al. 2010; Griep et al. 2013). During coordinated landing, the body is decelerated by the forelimbs and the landing forces are transmitted and absorbed by the forelimbs and the pectoral girdle (Emerson, 1983, 1984; Nauwelaerts & Aerts, 2006): The preparation for landing involves an extension of the elbow and a protraction and adduction of the humerus, whereas the opposite motions can be observed during landing (e.g., Griep et al. 2013; Gillis et al. 2014). At least the mm. anconaeus, coracoradialis,

episternohumeralis, scapulohumeralis superficialis (likely including m. cleidohumeralis in Akella & Gillis, 2011), and the portio sternalis of the m. pectoralis act to prepare the landing and to dissipate the impact forces during landing (Gillis et al. 2010; Akella & Gillis, 2011; muscle names according to the nomenclature suggested in Chapter six). The m. supracoracoideus (humerus adductor; Gaupp, 1896) has not been considered in Gillis et al. (2010) and Akella & Gillis (2011); given that this muscle is closely associated with the m. coracoradialis (see Chapter six and, e.g., Gaupp, 1896; Bigalke, 1927; Ritland, 1955), it seems likely that the activity of the m. supracoracoideus has been measured in conjunction with the activity of the m. coracoradialis in these previous studies. If so, the m. supracoracoideus would be active during landing and contribute to the dissipation of the landing forces, too.

The same mechanical principles that apply to the forces and lever arms involved in the retraction of the humerus also apply to the protraction and adduction of the humerus. Consequently, the efficiency of the humerus protractor muscles would be increased if the origin of those muscles was shifted to a more anterior position relative to the shoulder joint. Such a shift in the origin might be realized in some species by the presence of an episternum that provide areas for muscle attachments (compare, e.g., Trueb, 1973 and Chapter four) or an anteriorly expanded acromion. Likewise, the efficiency of the adductor muscles would be increased if the origin of those muscles was shifted to a more ventral position relative to the shoulder joint. Such a shift could, for example, be achieved by increasing the angle of the clavicula and coracoid to the horizontal plane. This, however, has neither been observed, nor explicitly tested for in Chapter four, as the landing behavior was not considered therein. In addition, the muscles involved in the humerus protraction and adduction would be most efficient if they inserted distally on the humerus and if the humerus was comparably short.

The above theoretical considerations on the biomechanics of the musculo-skeletal complex of the anuran shoulder joint highlight the importance of the shoulder joint muscle configuration for locomotion and the potential for adaptations to specific tasks. All the potential mechanisms to increase the efficiency of the humerus protractor and adductor muscles in dissipating the landing forces are theoretical and need to be supported by future studies that, for example, assess the geometry of the pectoral girdle skeleton, humerus, and respective muscles in species that show a coordinated lading behavior in comparison to species with less coordinated landing.

Comparison of the shoulder joint muscles in anurans and salamanders

The reconstruction of muscle character states in the last common ancestor of the Anura could substantially contribute to our understanding of anatomical correlates to the evolution of locomotor behaviors at the base of and within the Anura and of the evolution of the tetrapod limb (compare Chapters one and six). As the reconstruction of ancestral character states requires an appropriate outgroup, the sister group of the Anura, the Caudata (Pyron & Wiens, 2011; Kumar et al. 2017), would need to be considered. The knowledge of the shoulder joint muscle homologies between the Anura and the Caudata is, thus, of interest.

The descriptions of the shoulder joint muscles in *Salamandra salamandra* (Linnaeus, 1758) by Francis (1934) and unpublished data on the corresponding muscles in *Hynobius tokyoensis* Tago, 1931 (Fig. 3; nerves not observed) indicate that salamanders generally have fewer shoulder joint muscles than anurans. Generalizing the observations in *S. salamandra* and *H. tokyoensis*, reveals the following pattern of the shoulder joint muscle in salamanders (names of muscles and nerves follow Francis, 1934): The m. dorsalis scapulae covers most of the lateral surface of the suprascapula. This muscle originates from a long line near and parallel to the dorsal margin of the suprascapula and inserts with a strong tendon onto the anterior surface of the crista ventralis humeri, and is innervated by a branch of n. dorsalis scapulae. The posterior part of the m. dorsalis scapulae is covered by the m. dorso-humeralis, which originates from the fascia dorsalis and inserts via a strong common tendon with the m. anconaeus scapularis medialis with two attachments: (1) anterior surface of the crista ventralis humeri and (2) dorsal margin of the glenoid fossa. The m. dorso-humeralis is innervated by the n. dorso-humeralis. The m. anconaeus in salamanders has four head, two which originate from the pectoral girdle: The m. anconaeus scapularis medialis (also called ‘caput a’ in Francis, 1934) originates from by the common tendon with the m. dorso-humeralis from the dorsal margin of the glenoid fossa and the m. anconaeus coracoideus (also ‘caput b’) arises from the posterior margin and medial surface of the bony scapulocoracoid. All heads distally chare a common tendon that inserts onto the proximal part of ulna; the heads are innervated by the n. extensorius caudalis. Anteroventrally, the m. procoraco-humeralis originates from the ventral surface of the procoracoid cartilage and inserts onto the anterior surface of the crista ventralis humeri. This muscle is innervated by branches of the n. dorsalis scapula and the n. supracoracoideus.

A large part of the ventral surface of the cartilaginous coracoid is covered by the mm. coraco-radialis proprius and the supracoracoideus, with the former being covered by the latter. Both these muscles are continuous at their origins and arise from the ventral surface of the

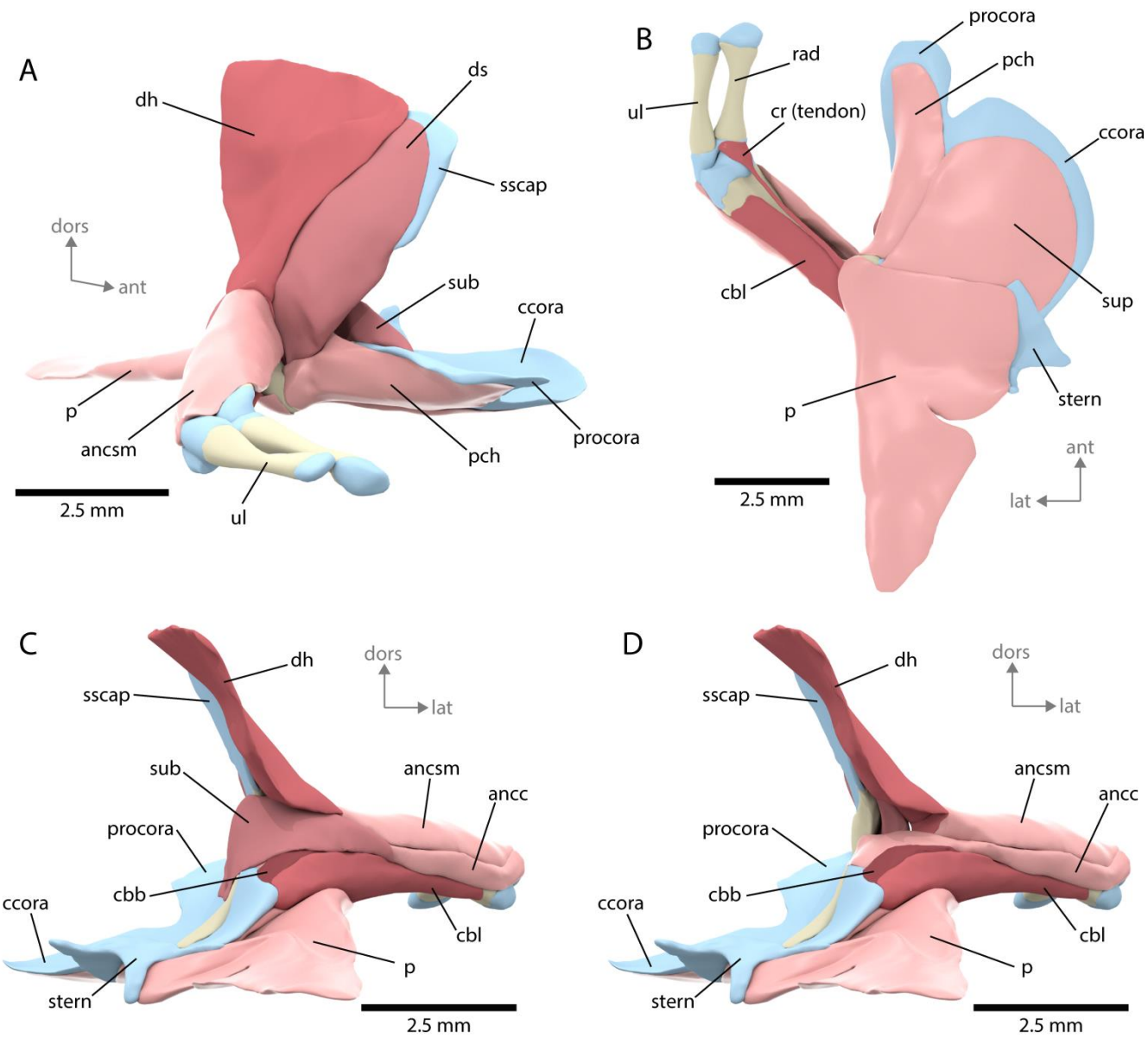


Figure 3 Shoulder joint muscles in *Hynobius tokyoensis* (ZMH A 12262). Surface model derived from episcopic images. Beige: bone; blue: cartilage; red: muscle (shades of red correspond to the homologous anuran muscles in Chapter six with the exception of lightest rose for muscles with no or uncertain omologous muscle in anurans). Muscle nomenclature follows Francis (1934). **A** Anterolaterl view. **B** Ventral view. **C** Posterior view. **D** As C, but m. subcoracoscapularis removed. anccc: m. anconaeus coracoideus; ancsm: m. anconaeus scapularis medialis; ant: anterior; cbb: m. coraco-brachialis brevis; cbl: m. coraco-brachialis longus; ccora: cartilaginous coracoid; cr: m. coraco-radialis proprius; dh: m. dorso-humeralis; dors: dorsal; ds: m. dorsalis scapulae; lat: lateral; pch: m. procoraco-humeralis; p: m. pectoralis; procora: procoracoid; rad: radius; sub: m. subscapularis; sscap: suprascapula; stern: sternum; sup: m. supracoracoideus; ul: ulna.

cartilaginous coracoid. The m. supracoracoideus inserts onto the ventral part of the posterior surface of the crista ventralis humeri. This muscle is continuous with the m. pectoralis at its insertion and is innervated by the n. supracoracoideus. The m. coraco-radialis proprius laterally forms a strong tendon with two insertions: (1) the proximal head of the humerus near the insertion of the m. supracoracoideus and (2) the proximal part of radius. This muscle

Table 1 Hypothesis on the homologies of the shoulder joint muscles in anurans and salamanders. Names for muscles in anurans follow the suggested terminology in Chapter six; the names for salamander muscles follow Francis (1934). Uncertain homologizations are indicated by a question mark (?).

Muscle(s) in Anura	**Muscle(s) in Caudata**
m. dorsalis scapulae	m. dorsalis scapula
m. latissimus dorsi	m. dosro-humeralis
? m. anconaeus caput scapulare	? mm. anconaeus coracoideus and/or anconaeus scapularis medialis
? mm. scapulohumeralis superficialis, scapulohumeralis profundus anterior, scapulohumeralis profundus posterior, cleidohumeralis, and/or episternohumeralis	? m. procoraco-humeralis
m. supracoracoideus and ? m. episternohumeralis	m. supracoracoideus
m. coracoradialis	m. coraco-radialis proprius
m. pectoralis portionis sternalis and abdominalis	m. pectoralis
m. pectoralis portio coracoidea	m. coraco-brachialis longus
m. coracobrachialis	m. coraco-brachialis brevis
m. subcoracoscapularis	m. subscapularis

receives supply from the n. supracoracoideus. The anterior fibers of the m. pectoralis arise from an aponeurosis that separates this muscle from its counterpart of other body side, then, the origin continues along the ventral surface of the sternum and the most posterior fibers arise from the ventral surface of the m. rectus abdominis superficialis. The m. pectoralis inserts on the distal part of posterior surface of crista ventralis humeri and is innervated by the n. pectoralis.

The shoulder joint is posteriorly covered by the mm. coraco-brachialis longus et brevis. The m. coraco-brachialis longus originates from the posterolateral margin of the cartilaginous coracoid and inserts laterally onto the posteroventral surface of the humerus. The m. coraco-brachialis brevis arises from the posterior part of the ventral surface of the scapulocoracoid and adjacent cartilaginous coracoid. This muscle inserts onto the posterior surface of the humerus and its crista ventralis. Both coraco-brachialis muscles are innervated by branches (nn. coraco-brachialis) of the r. superficialis of the n. brachialis. In addition, the n. brachialis passes between these two muscles. The m. subscapularis originates from the dorsal surface of the procoracoid and inserts onto the posterior surface of the humerus or the crista ventralis. It is supplied by the n. subscapularis.

The shoulder joint muscles in anurans and salamanders can be homologized as hypothesized in Table 1 by applying the criterions for primary homology used in Chapter six (also see Remane, 1952; de Pinna, 1991). There, however, remain some uncertain homologies: In the Anura, only one of the heads, the caput scapulare, of the m. anconaeus originates from the pectoral girdle, whereas there are two such heads in salamanders; neither the relative position of these heads nor their innervation provide enough evidence to

hypothesize which of the anconaeus heads (or both) in salamanders is homologous to the caput scapulare in anurans. The m. procoraco-humeralis in salamanders claims about the same space as the group of scapulohumeralis muscles and the mm. cleidohumeralis and episternohumerlais in anurans. The innervation of the m. procoraco-humeralis by the nn. dorsalis scapula and supracoracoideus (likely homologous to the n. dorsalis scapulae anterior and supracoracoideus in anurans) salamanders also corresponds to the innervation of the possibly homologous muscle in anurans. These anuran muscles, however, are ontogenetically derived from different pre-muscle masses (see Chapter six), which leads to several possible hypothesis on the inter-order homology of these muscles. For example, the m. procoraco-humeralis in salamanders may consist of fibers that are ontogenetically derived from the ventral and the dorsal pre-muscle mass; if so, the procoraco-humeralis might be homologous with all the named muscle in anurans. Alternatively, the m. procoraco-humeralis might ontogenetically be derived from either the dorsal or the ventral pre-muscle mass. In the former case the m. procoraco-humeralis could be homologous with some or all scapulohumeralis muscles in anurans, in the latter case it could be homologous with the mm. cleidohumeralis and episternohumeralis. Yet, the observations in Chapter six show that the m. episternohumeralis is closely related with the m. supracoracoideus in some anuran species; this allows the hypothesis, that the anuran m. episternohumeralis might be a part of the m. supracoracoideus in salamanders. Studies on the ontogenetic origin of the shoulder joint muscles in salamanders are needed to derive better supported hypothesis on the homology of the shoulder joint muscles in anurans and salamander.

Morphological methods

The results reported in the Chapters two and three demonstrated that bones can be segmented with high accuracy in micro-computed tomography (CT) volumes and that surfaces with sub-voxel accuracy can be derived from the segmentation results if suitable segmentation and surface generation approaches are applied. This gives confidence that µCT is a valid approach to acquire quantitative measurements of bones and likely other tissues that show sufficient contrast to adjacent structures.

Concerning other tissues than bone, an informal literature review of published studies that performed digital dissections of µCT volumes of vertebrates revealed differences in the anatomical accuracy and quality of surface models used to illustrate the anatomy of the respective specimens. In several studies, cartilage has not been visualized (particularly

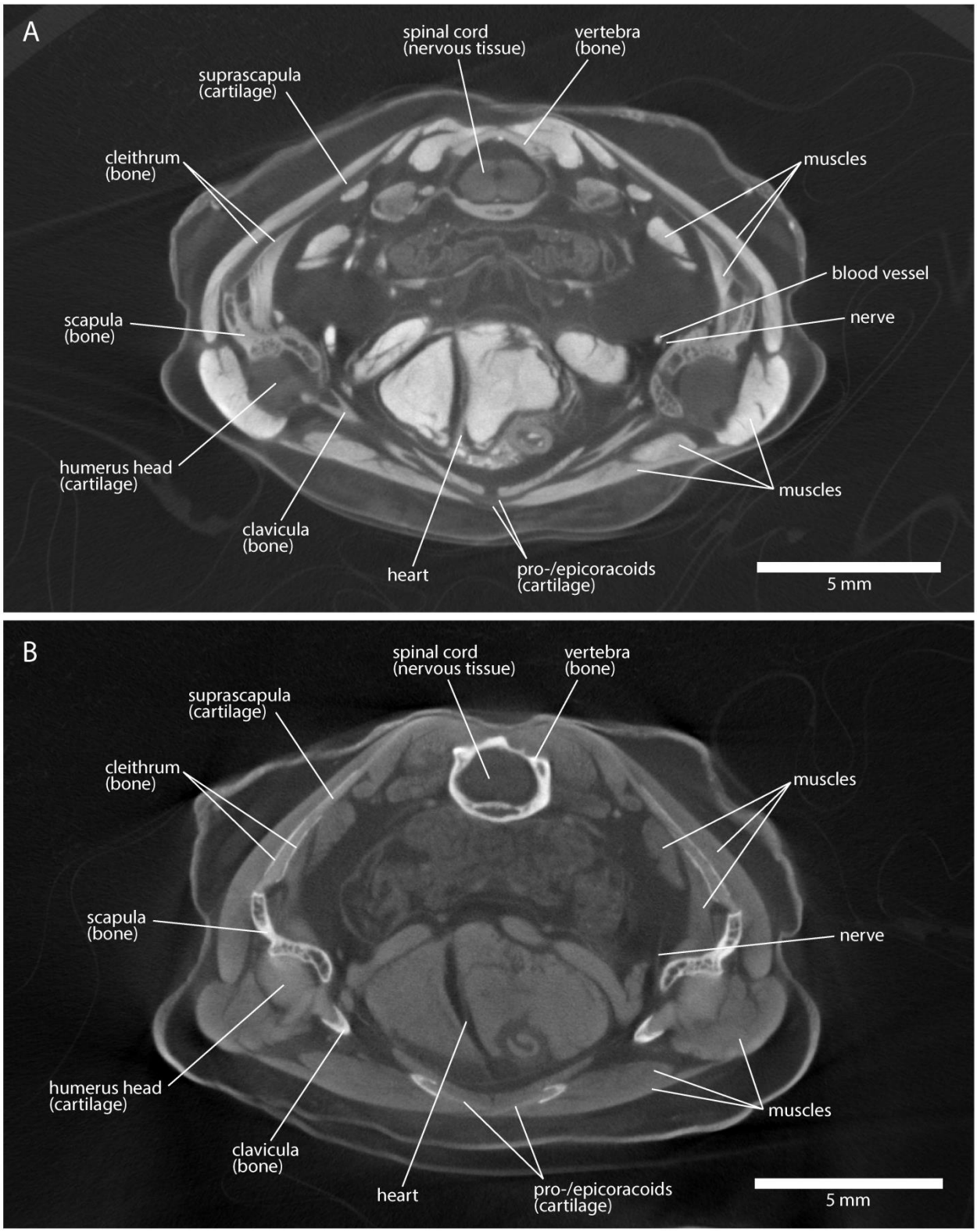


Figure 4 Cross sections of contrast enhanced µCT volumes of *Pelodytes punctatus* specimens showing corresponding parts of the pectoral girdle region. (A) Iodine-stained specimen ZMH A07271. (B) Lead-stained specimen ZMH A07240 (both CT scans performed by Angelika Ziolkowski).

obvious in joints where the corresponding bones are ending without connection) and the texts rarely provided a notion or discussion on this shortcoming (e.g., Klinkhamer et al. 2017;

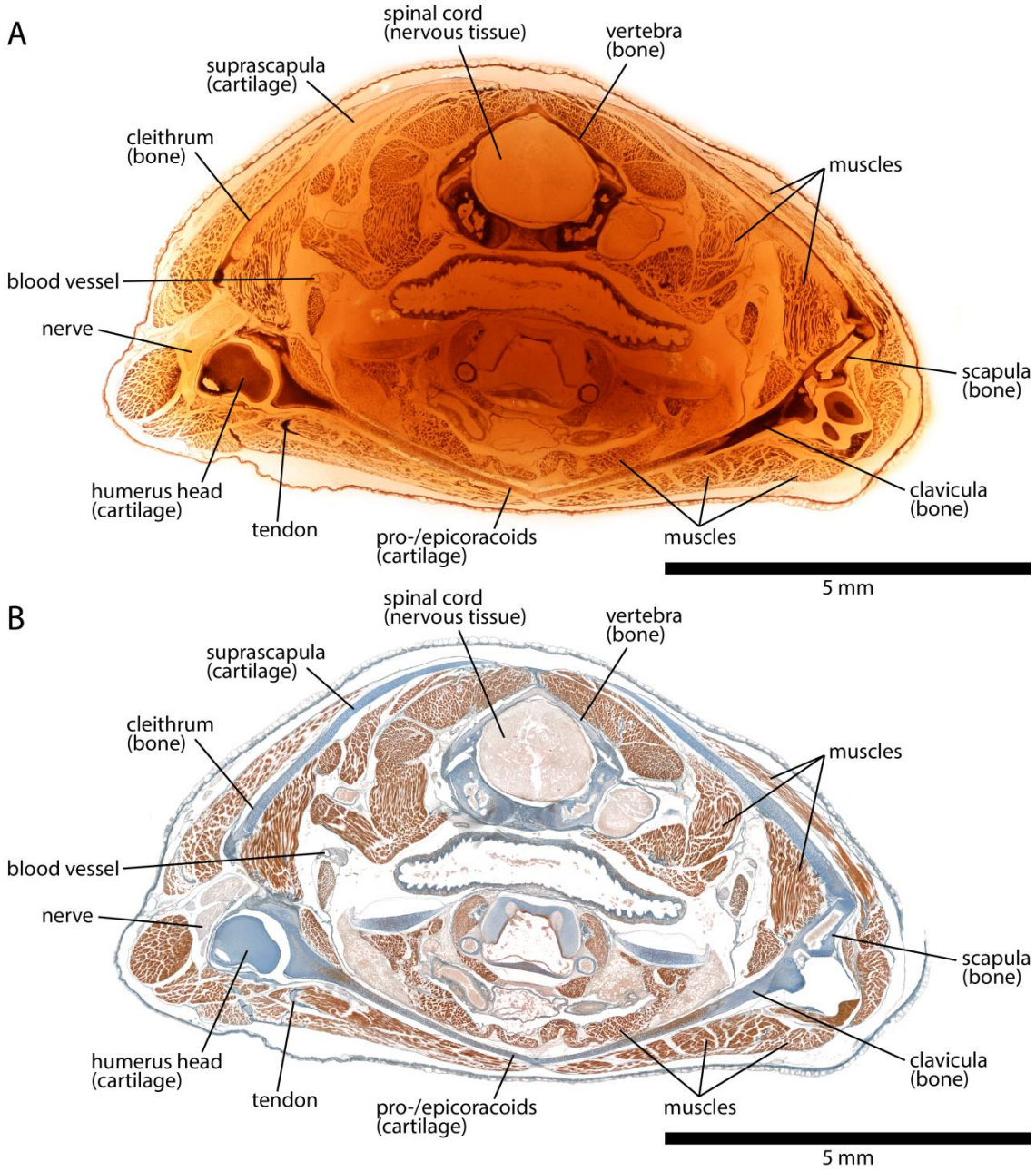


Figure 5 Comparison of corresponding (A) sodium sulfide stained episcopic and (B) Azan stained histological sections of the pectoral girdle region of *Xenopus laevis* (ZMH A05087) (raw images generated in Master project of Juliana Lutz, 2018).

Porro & Richards, 2017; Bellati et al. 2018; Collings & Richards, 2019). Some surfaces of skeletal elements showed uncommented artefactual holes that might be mistaken as foramina, or, conversely, foramina that were erroneously closed (e.g., Krings et al. 2017; Porro & Richards, 2017). Other skeletal elements have been represented by one single and continuous surfaces, whereas there should have been free spaces or distinct surface elements illustrating the separation of structures like adjacent skeletal elements (e.g., Krings et al. 2017) or jaws and teeth (e.g., Holliday et al. 2013). Some surfaces of muscles have freely ended without connections to any structures (e.g., Porro & Richards, 2017; Brocklehurst et al. 2019; Collings

& Richards, 2019) or have comprised artefactual holes or irregularities that might be mistaken as entrances of nerves or blood vessels (e.g., Krings et al. 2017; Porro & Richards, 2017). Illustrations of discontinuous blood vessels could also been found (Weinhardt et al. 2018).

Such inaccuracies in surface models might be the consequence of the poor visibility or invisibility of some tissue like nerves, tendons, blood vessels, and uncalcified cartilage in (contrast-enhanced) µCT volumes (personal observation; Fig. 4; Lautenschlager et al. 2014; Bribiesca-Contreras & Sellers, 2017; Sullivan et al. 2019). However, these tissues have successfully been visualized in detail in contrast-enhanced µCT scans under specific conditions (tendon: Shearer et al. 2014; Sartori et al. 2018; nerves/neurons: de Castro Fonseca et al. 2018; Töpperwien et al. 2018; blood vessels: Porter & Witmer, 2015; Qiu et al. 2016). Resolutions and contrasts comparable to histological sections (Busse et al. 2018) or the visualization of cell nuclei (Müller et al. 2018) have been achieved with µCT if specimen preparation (including contrast staining with appropriately chosen agents) and scanning conditions were appropriate. Most of the referred examples of successful visualizations of the sometimes poorly visible tissues have in common that considerable effort was spent on the specimen preparation including, for example trimming the specimen to the anatomical structures of interest or trials on determining the optimal staining agent and protocol.

At some point it might be necessary to consider alternative methodological approaches to generate volume data for digital dissections, particularly if the entire specimen and not only parts of it should be imaged in detail. Episcopic microtomy (Chapter five; Fig. 5A) and properly stained histological serial sectioning (Fig. 5B) are potential alternatives to µCT. These methods offer higher resolutions in the plane of sectioning and higher tissue contrasts. The different methods of volume data acquisition can be combined. A lead-impregnated and paraffin embedded specimen could, for example, be µCT scanned and subsequently imaged episcopically; selected sections could be mounted on glass slides and histologically stained. This would realize the benefits of the different approaches (also see, e.g., Handschuh et al. 2013; Herdina et al. 2015).

Importance of accuracy of illustrations

Illustrations have been used to complement textual descriptions for conveying anatomical observations (e.g. Carmichael & Pawlina, 2000; Farrell, 2006; Ghosh, 2015). Illustrations can even outperform written descriptions when it comes to communicating anatomical knowledge from doctors to patients in a medical context (McGhee, 2010, and references therein). Vice

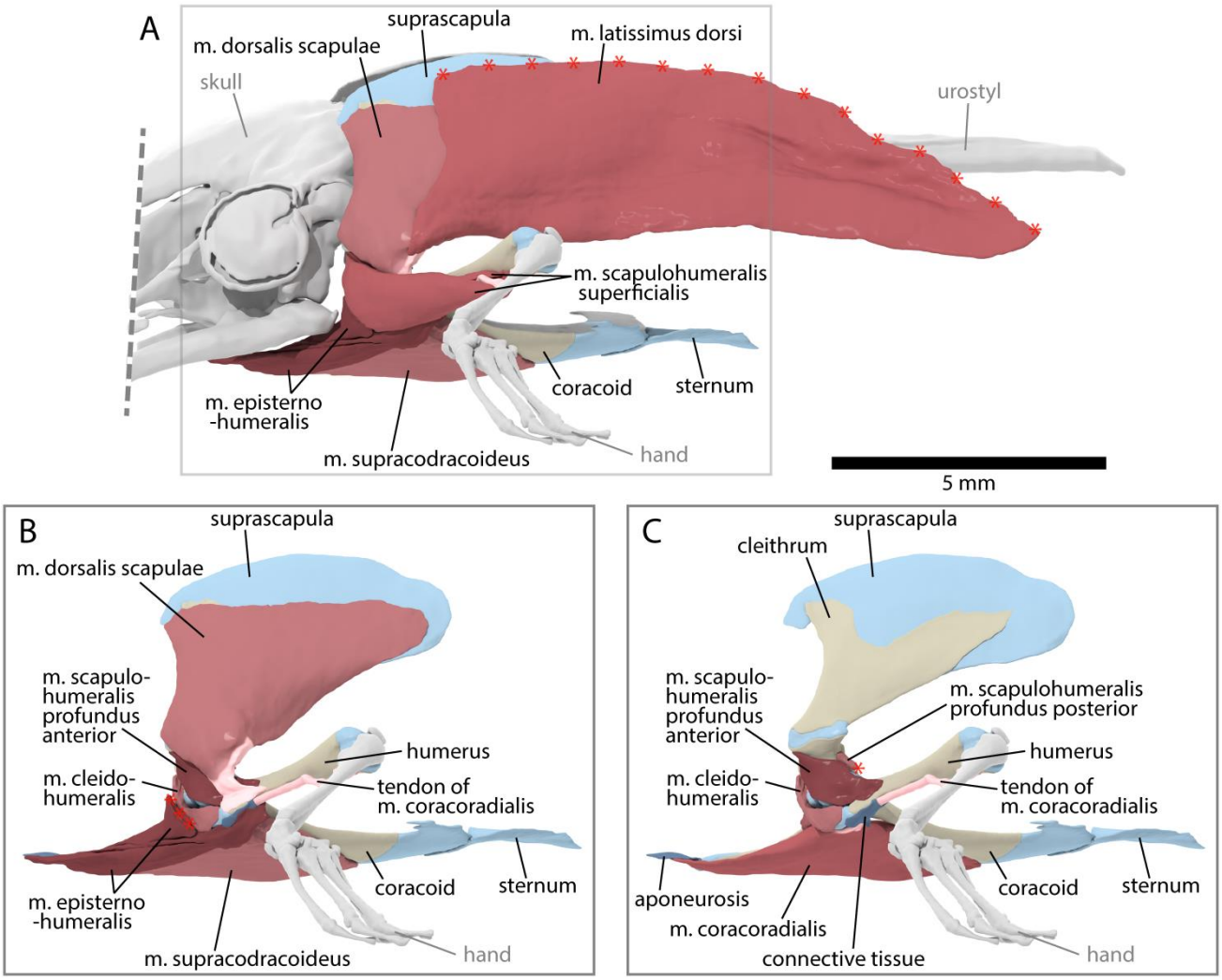


Figure 6 Selected muscles and contextual skeletal elements of the left side pectoral girdle of *Xenopus laevis* (ZMH A05087) in lateral view; anterior to the left. Surface model derived from volume of episcopic images. Red: muscle; rose: tendon; beige: bone; light blue: cartilage; dark blue: aponeurosis/connective tissue; gray: skeletal structures without distinguishing between elements and bone and cartilage, and neglecting some anatomical details. Dotted line: underlying element cut at this position. Red asterisk: element of origin/insertion for muscle not illustrated. (A) Superficial layer of selected muscles. (B) Detail of (A) with skull, right pectoral girdle half, urostyl, and mm. latissimus dorsi and scapulohumeralis superficialis removed. (C) Same as (B) but mm. episternohumeralis, supracoracoideus, and dorsalis scapulae removed (surface model modified from Lutz, 2018 [Master Thesis]).

versa, erroneous anatomical illustrations might rather confuse than help readers who aim to gain anatomical knowledge from illustrated texts (Brödel, 1907, 1941).

Within the general field of education, Mishra (1999) concluded that all "forms of illustration depend on artistic conventions – conventions are not 'natural' and have to be learnt" (Mishra, 1999, p. 156). Transferring this insight into the field of biology, readers are accustomed to accurate anatomical illustrations that depict a rather complete anatomical context, even if some of the illustrated elements are not of importance for what the author

intends to communicate. For example, illustrations of muscles traditionally show the contextual bones and cartilaginous elements, and cut and removed muscles often are still judiciously depicted at their origins and insertions to provide structural reference and connectivity (compare, e.g. Gaupp, 1896; Bigalke, 1927; Ritland, 1955). If illustrations of µCT volumes or derived surfaces neglect (parts of) structures or contain artefacts, readers unfamiliar with such illustrations might not recognize these inaccuracies. As a consequence, the readers might extract incomplete or wrong anatomical information from the illustrations.

The literature reports above highlight the importance of illustrations for communicating anatomical knowledge. They also stress that illustrations need to be complete and factually correct, which also applies to surface models derived from digital dissections. If artefacts or the omission of structures due to poor visibility are unavoidable, this needs to be clearly communicated, for example, by noting them in the figure caption or by directly indicating them in the illustration. Missing structures or cut structures (e.g., parts of a bone outside the µCT scanner's field of view) should also be clearly emphasized; a dotted line, for example, could indicate the position at which an element is cut or a figure caption could state "bones of specimen X" instead of "skeleton of specimen X" if cartilage wasomitted (see Fig. 6 for some more examples). This will help ensuring a clear and accurate communication of anatomical observations.

References

Akella T, Gillis GB (2011) Hopping isn't always about the legs: forelimb muscle activity patterns during toad locomotion. J Exp Zool A Ecol Genet Physiol **315**, 1–11.

Bellati A, Scherz MD, Megson S, et al. (2018) Resurrection and re-description of *Plethodontohyla laevis* (Boettger, 1913) and transfer of *Rhombophryne alluaudi* (Mocquard, 1901) to the genus *Plethodontohyla* (Amphibia, Microhylidae, Cophylinae). ZSE **94**, 109–135.

Bigalke R (1927) Zur Myologie der Erdkröte. (*Bufo vulgaris*, Laurenti.). Z Anat Entwickl Gesch **82**, 286–353.

Bribiesca-Contreras F, Sellers WI (2017) Three-dimensional visualisation of the internal anatomy of the sparrowhawk (*Accipiter nisus*) forelimb using contrast-enhanced micro-computed tomography. PeerJ **5**, e3039.

Brocklehurst R, Porro L, Herrel A, Adriaens D, Rayfield E (2019) A digital dissection of two teleost fishes: comparative functional anatomy of the cranial musculoskeletal system in pike (*Esox lucius*) and eel (*Anguilla anguilla*). J Anat **235**, 189–204.

Brödel M (1907) How may our present methods of medical illustration be improved? JAMA **XLIX**, 138.

Brödel M (1941) Medical illustration. JAMA **117**, 668.

Busse M, Müller M, Kimm MA, et al. (2018) Three-dimensional virtual histology enabled through cytoplasm-specific X-ray stain for microscopic and nanoscopic computed tomography. Proc Natl Acad Sci U.S.A. **115**, 2293–2298.

Carmichael SW, Pawlina W (2000) Animated PowerPoint as a tool to teach anatomy. Anat Rec **261**, 83–88.

Collings AJ, Richards CT (2019) Digital dissection of the pelvis and hindlimb of the red-legged running frog, *Phlyctimantis maculatus*, using diffusible iodine contrast enhanced computed microtomography (DICE μCT). PeerJ **7**, e7003.

Davidovits P (2008) Physics in biology and medicine. Amsterdam, Boston: Elsevier/Academic Press. Available from: http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=214762.

Davis CM (1974) The role of effective lever length in the perception of lifted weights. Perception & Psychophysics **16**, 67–69.

de Castro Fonseca M, Araujo BHS, Dias CSB, et al. (2018) High-resolution synchrotron-based X-ray microtomography as a tool to unveil the three-dimensional neuronal architecture of the brain. Sci Rep **8**, 12074.

de Pinna MCC (1991) Concepts and tests of homology in the cladistic paradigm. Cladistics **7**, 367–394.

Emerson SB (1976) Burrowing in frogs. J Morphol **149**, 437–458.

Emerson SB (1983) Functional analysis of frog pectoral girdles. The epicoracoid cartilages. J Zool **201**, 293–308.

Emerson SB (1984) Morphological variation in frog pectoral girdles: testing alternatives to a traditional adaptive explanation. Evolution **38**, 376–388.

Emerson SB (1991) A biomechanical perspective on the use of forelimb length as a measure of sexual selection in frogs. J Evolution Biol **4**, 671–678.

Essner RL, Suffian DJ, Bishop PJ, Reilly SM (2010) Landing in basal frogs: evidence of saltational patterns in the evolution of anuran locomotion. Naturwissenschaften **97**, 935–939.

Farrell A (2006) Getting the picture: medical images on the web. Med Ref Serv Q **25**, 47–54.

Francis ETB (1934) The anatomy of the Salamander. London: Oxford University Press.

Gaupp EWT (1896) A. Ecker's und R. Wiedersheim's Anatomie des Frosches. Erste Abteilung. Lehre vom Skelett und vom Muskelsystem. Braunschweig: Friedrich Vieweg und Sohn.

Ghosh SK (2015) Evolution of illustrations in anatomy: a study from the classical period in Europe to modern times. Anat Sci Educ **8**, 175–188.

Gillis GB, Akella T, Gunaratne R (2010) Do toads have a jump on how far they hop? Pre-landing activity timing and intensity in forelimb muscles of hopping Bufo marinus. Biol Lett **6**, 486–489.

Gillis GB, Ekstrom L, Azizi E (2014) Biomechanics and control of landing in toads. Integr Comp Biol **54**, 1136–1147.

Griep S, Schilling N, Marshall P, Amling M, Hahne LM, Haas A (2013) Pectoral girdle movements and the role of the glenohumeral joint during landing in the toad, *Rhinella marina* (Linnaeus, 1758). Zoomorphology **132**, 325–338.

Handschuh S, Baeumler N, Schwaha T, Ruthensteiner B (2013) A correlative approach for combining microCT, light and transmission electron microscopy in a single 3D scenario. Front Zool **10**, 44.

Helfsgott J (2020) Die Beziehung zwischen der Form des Humerus und Lokomotionsmodi bei Froschlurchen – eine geometrisch-morphometrische Analyse. Master Thesis. Hamburg.

Herdina AN, Plenk H, Benda P, et al. (2015) Correlative 3D-imaging of *Pipistrellus* penis micromorphology: validating quantitative microCT images with undecalcified serial ground section histomorphology. J Morphol **276**, 695–706.

Hildebrand M (1995) Analysis of vertebrate structure. New York: Wiley. Available from: http://www.loc.gov/catdir/enhancements/fy0608/94004652-d.html.

Holliday CM, Tsai HP, Skiljan RJ, George ID, Pathan S, Farke AA (2013) A 3D interactive model and atlas of the jaw musculature of *Alligator mississippiensis*. PLoS ONE **8**, e62806.

Keeffe R, Blackburn DC (2020) Comparative morphology of the humerus in forward-burrowing frogs. Biol J Linnean Soc **171**, 293.

Klinkhamer AJ, Wilhite DR, White MA, Wroe S (2017) Digital dissection and three-dimensional interactive models of limb musculature in the Australian estuarine crocodile (*Crocodylus porosus*). PLoS ONE **12**, e0175079.

Krings M, Klein B, Heneka MJ, Rödder D, Lötters S (2017) Morphological comparison of five species of poison dart frogs of the genus *Ranitomeya* (Anura: Dendrobatidae) including the skeleton, the muscle system and inner organs. PLoS ONE **12**, e0171669.

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Molecular biology and evolution **34**, 1812–1819.

Lautenschlager S, Bright JA, Rayfield EJ (2014) Digital dissection – using contrast-enhanced computed tomography scanning to elucidate hard- and soft-tissue anatomy in the Common Buzzard *Buteo buteo*. J Anat **224**, 412–431.

Lieber RL, Fridn J (2000) Functional and clinical significance of skeletal muscle architecture. Muscle Nerve **23**, 1647–1666.

Lutz J (2018) Die Schultergürtelmorphologie von *Xenopus laevis* (Daudin, 1802) mit Hinweisen zu Ontogenese und zwischenartlicher Muskelhomologie. Master Thesis. Hamburg.

McGhee J (2010) 3-D visualization and animation technologies in anatomical imaging. J Anat **216**, 264–270.

Mishra PB (1999) The role of abstraction in scientific illustration: implications for pedagogy. J Vis Lit **19**, 139–158.

Müller M, Kimm MA, Ferstl S, et al. (2018) Nucleus-specific X-ray stain for 3D virtual histology. Sci Rep **8**, 17855.

Nauwelaerts S, Aerts P (2006) Take-off and landing forces in jumping frogs. J Exp Biol **209**, 66–77.

Nomura F, Rossa-Feres DC, Langeani F (2009) Burrowing behavior of *Dermatonotus muelleri* (Anura, Microhylidae) with reference to the origin of the burrowing behavior of Anura. J Ethol **27**, 195–201.

Porro LB, Richards CT (2017) Digital dissection of the model organism *Xenopus laevis* using contrast-enhanced computed tomography. J Anat **207**, 399.

Porter WR, Witmer LM (2015) Vascular patterns in iguanas and other squamates: blood vessels and sites of thermal exchange. PLoS ONE **10**, e0139215.

Přikryl T, Aerts P, Havelková P, Herrel A, Roček Z (2009) Pelvic and thigh musculature in frogs (Anura) and origin of anuran jumping locomotion. J Anat **214**, 100–139.

Pyron AR, Wiens JJ (2011) A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol Phylogenet Evol **61**, 543–583.

Qiu X, Shi X, Ouyang J, Xu D, Zhao D (2016) A method to quantify and visualize femoral head intraosseous arteries by micro-CT. J Anat **229**, 326–333.

R Core Team (2019) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Reilly SM, Jorgensen ME (2011) The evolution of jumping in frogs: morphological evidence for the basal anuran locomotor condition and the radiation of locomotor systems in crown group anurans. J Morphol **272**, 149–168.

Remane A (1952) Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Theorerische Morphologie und Systematik I. Leipzig: Akademische Verlagsgesellschaft Geest & Portig.

Ritland RM (1955) Studies on the post-cranial morphology of *Ascaphus truei*. II. Myology. J Morphol **97**, 215–282.

RStudio Team (2018) RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Available from: http://www.rstudio.com/.

Sartori J, Köhring S, Witte H, Fischer MS, Löffler M (2018) Three-dimensional imaging of the fibrous microstructure of Achilles tendon entheses in *Mus musculus*. J Anat **233**, 370–380.

Shearer T, Rawson S, Castro SJ, et al. (2014) X-ray computed tomography of the anterior cruciate ligament and patellar tendon. Muscles Ligaments Tendons J **4**, 238–244.

Stepanova N, Womack MC (2020) Anuran limbs reflect microhabitat and distal, later-developing bones are more evolutionarily labile. Evolution; international journal of organic evolution.

Sullivan SP, McGechie FR, Middleton KM, Holliday CM (2019) 3D muscle architecture of the pectoral muscles of European Starling (*Sturnus vulgaris*). Integr Org Biol **1**, oby010.

Töpperwien M, van der Meer F, Stadelmann C, Salditt T (2018) Three-dimensional virtual histology of human cerebellum by X-ray phase-contrast tomography. Proc Natl Acad Sci U.S.A. **115**, 6940–6945.

Trueb L (1973) Bones, frogs, and evolution. In *Evolutionary biology of the anurans. Contemporary research on major problems* (ed Vial JL), pp. 65–132. Columbia: University of Missouri Press.

Wainwright PC, Alfaro ME, Bolnick DI, Hulsey CD (2005) Many-to-one mapping of form to function: a general principle in organismal design? Integr Comp Biol **45**, 256–262.

Weinhardt V, Shkarin R, Wernet T, Wittbrodt J, Baumbach T, Loosli F (2018) Quantitative morphometric analysis of adult teleost fish by X-ray computed tomography. Sci Rep **8**, 16531.

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen und mich in den letzten Jahren unterstützt und begleitet haben!

Zuerst gilt mein Dank Herrn Prof. Dr. Alexander Haas für die Vergabe des Themas, die Begutachtung dieser Arbeit und die Unterstützung.

Auch Herrn Prof. Dr. Andreas Schmidt-Rhaesa danke ich für die Begutachtung dieser Arbeit.

Mein ganz besonderer Dank gilt Herrn Dr. Frank Friedrich, Dr. Jörg U. Hammel und Dr. Thomas Kleinteich. Herzlichen Dank für die vielen interessanten Diskussionen, die vielfältige Unterstützung auf ganz verschiedenen Ebenen, den Spaß an gemeinsamen Projekten, eure Geduld mit mir und für alles, was ich von euch lernen durfte! Ihr habt die letzten Jahre zu etwas ganz Besonderem für mich gemacht!

Außerdem bedanke ich mich ganz herzlich bei Dr. André Beerlink und bei der YXLON International GmbH für die tollen CT Scans, die spannenden Einblicke und die Erfahrungen, die ich in den gemeinsamen Projekten sammeln durfte!

Anna Maria Vogt, Lena Schwinger, Katharina Gebauer, Cordula Bracker, Lisa Gottschlich und Markus Dams danke ich aufrichtig für ihre tatkräftig und wertvolle Unterstützung im technischen Bereich und für die vielen netten und aufmunternden Gespräche.

Ebenso gilt mein besonderer Dank Juliana Lutz, Jennice Helfsgott, Mehria Sedik, Supak Panpeng, Sophie Schwippert und Vivienne Ahrens für die zahlreichen Plaudereien, die fachlichen Diskussionen, die Arbeit an spannenden Projekten und für alles, was ich dabei lernen konnte. Auch Lena Kath, Carla Vio Reinbold, Dilara Leptin, Laura Scherer und Angelika Ziolkowski danke ich ganz herzlich für ihre Unterstützung in verschiedenen Bereichen dieser Arbeit. Es hat mir sehr viel Spaß gemacht mich euch allen zu arbeiten!

Außerdem bin ich der gesamten AG Herpetologie (insbesondere Dr. Jakob Hallermann, Dr. Umilaela Arifin und André Jankowski), allen Kolleginnen und Kollegen am CeNak, den ITlern, dem Pförtner-Team und der Haustechnik sehr dankbar für die kleinen und großen Unterstützungen in ganz verschiedenen Bereichen.

Abseits des wissenschaftlichen Bereichs danke ich Carolin Vietz, Anika Hoffmann, Nina Pfeiffer, Svenja Hellemann, Christina Feldhaus und Lea Meisner von ganzem Herzen. Ihr hattet immer ein offenes Ohr für mich und habt mir eine ganze Menge Kraft gegeben! Tausend Dank dafür! Ohne euch gäbe es diese Arbeit sicher nicht.

Unendlich dankbar bin ich meiner Familie: Meinen Eltern, meiner Schwester und meinen Großeltern! Ihr habt mich mein ganzes Leben lang bedingungslos und auf alle erdenklichen Weisen unterstützt und mich in schweren Zeiten aufgebaut! Damit habt ihr mir den Rückhalt und die Kraft gegeben, ohne die ich nicht durch das Studium und die Zeit als Doktorandin gekommen wäre. Meine Dankbarkeit dafür kann ich gar nicht in Worten ausdrücken!

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, den 21.09.2020, 

Karolin Engelkes

Supporting Information

Supporting Information Chapter two

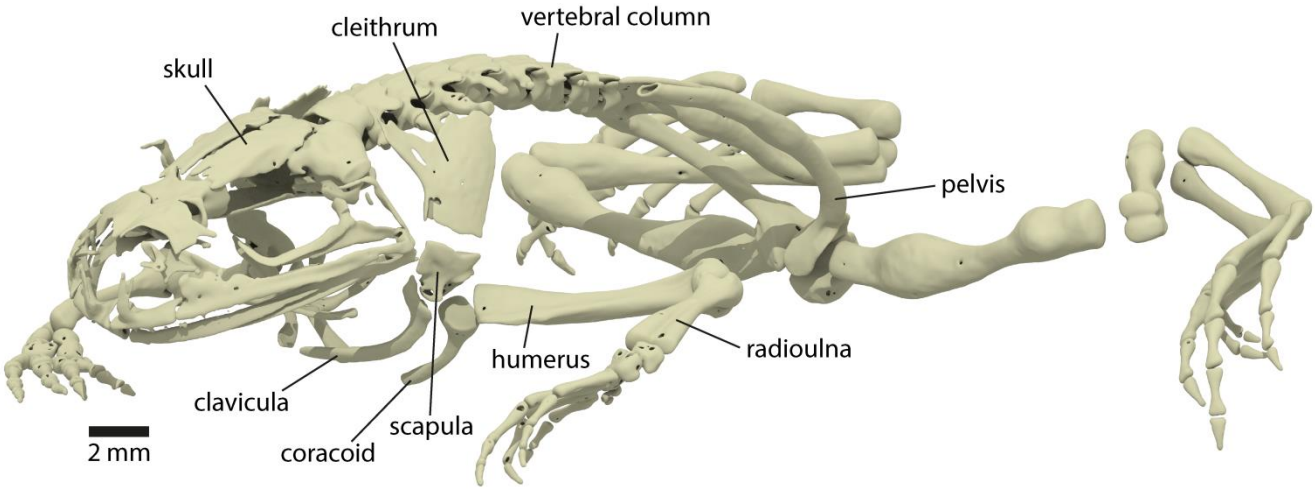


Fig. S1 Bones of *Bombina orientalis* (ZMH A12601). Surface render in anterolateral view, anterior to the left.

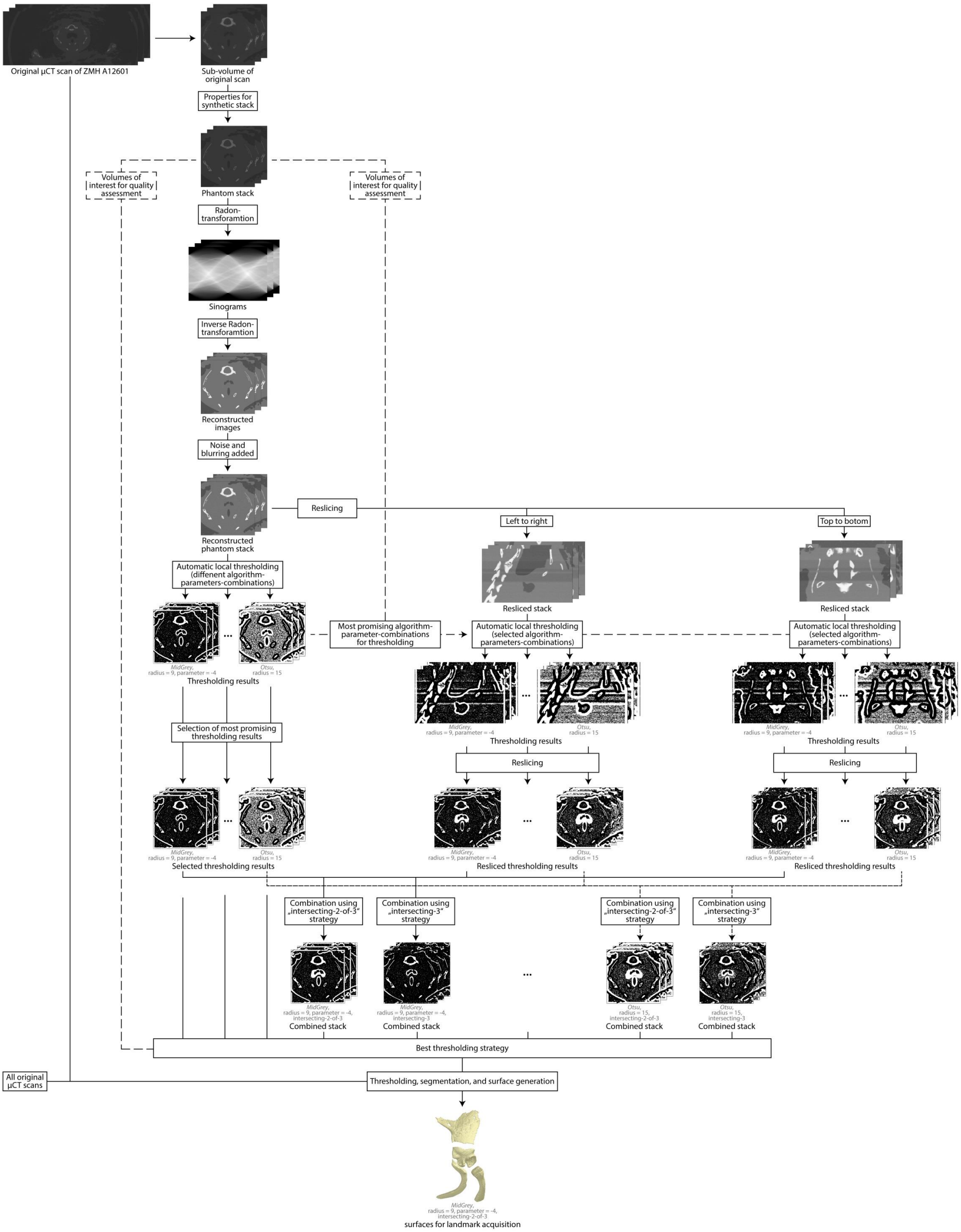
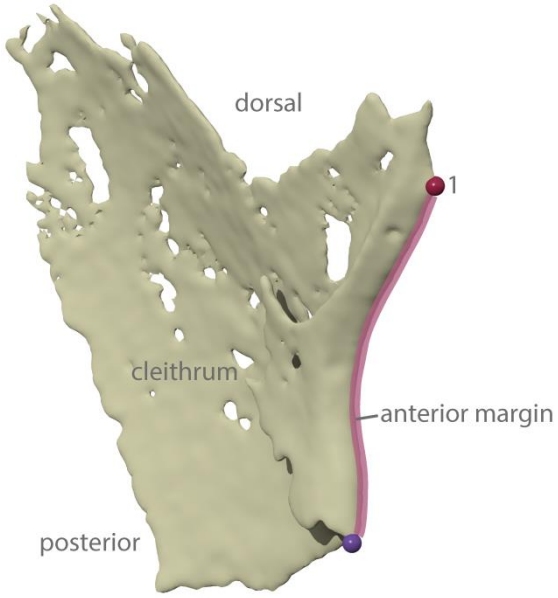
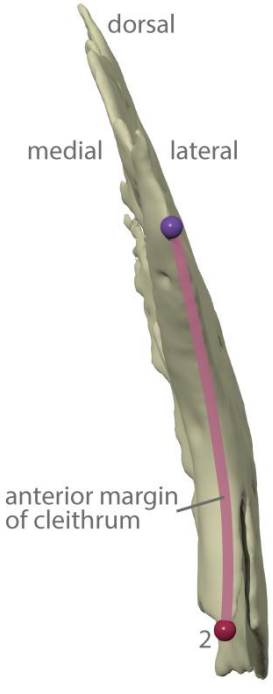
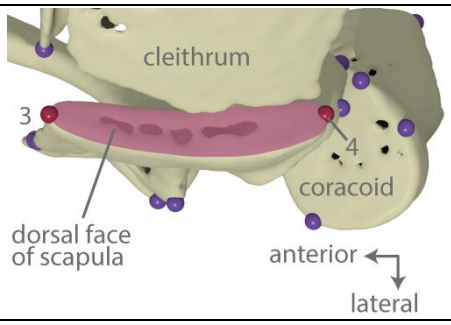


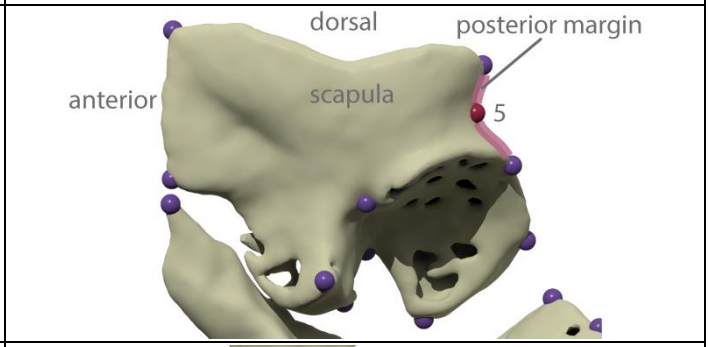
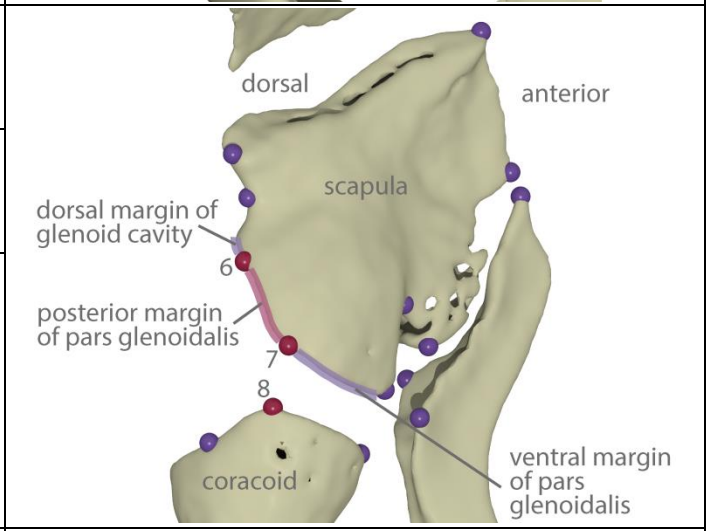
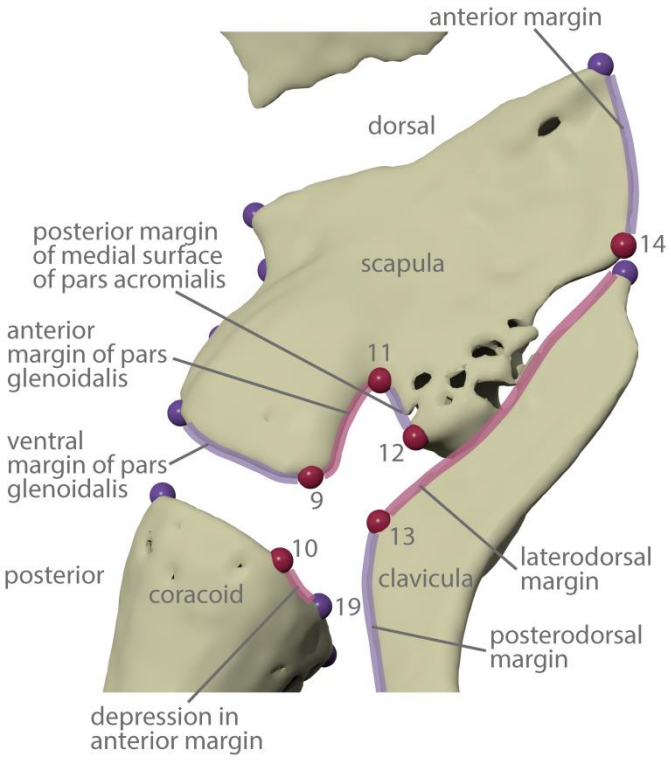
Fig. S2 Steps to determine the optimal segmentation strategy for the µCT scan of a selected *Bombina orientalis* specimen (ZMH A12601) using automatic local thresholding.

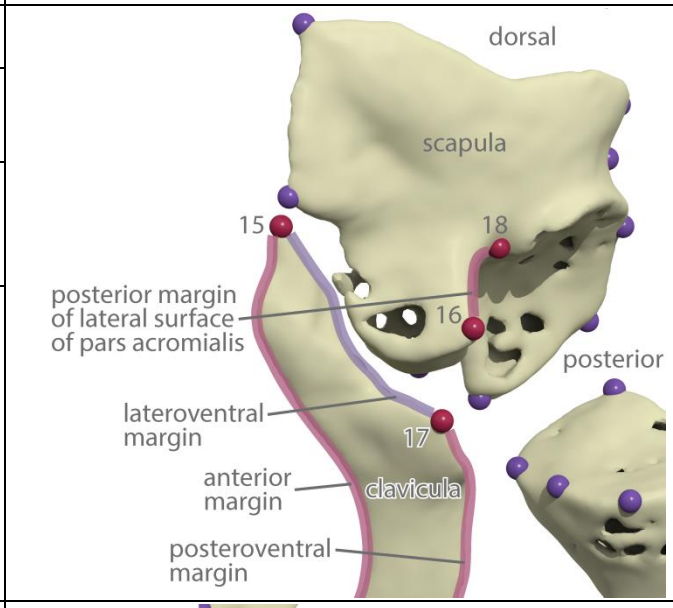
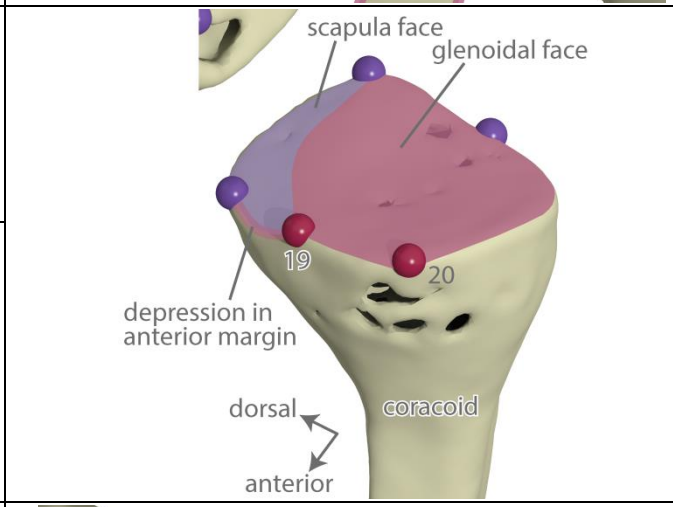
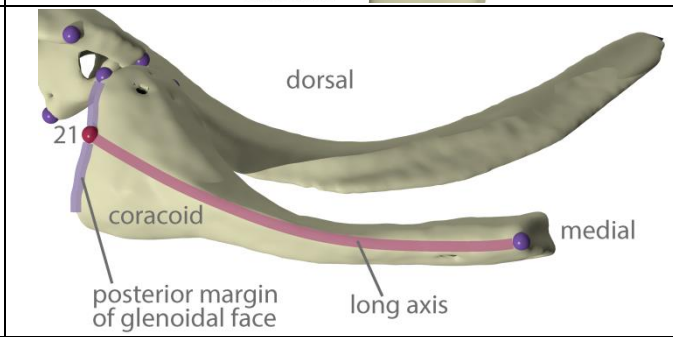
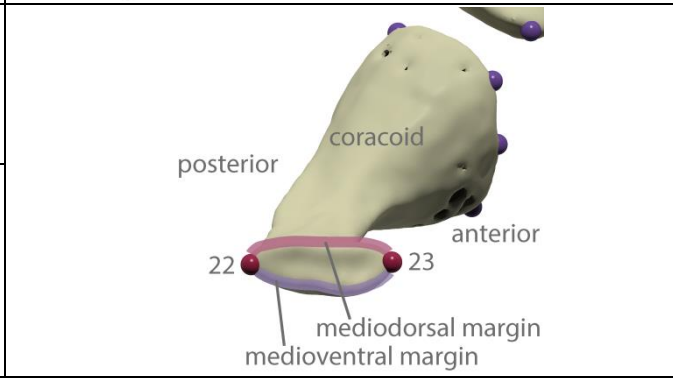
Supporting Information Table S1 Thresholding quality assessed for bone volumes and the adjacent two rows of voxels (first quality measure) of automatic local thresholding trials. The best and second best algorithm-parameter-combinations were performed on the reconstructed phantom stack and its resliced derivatives.

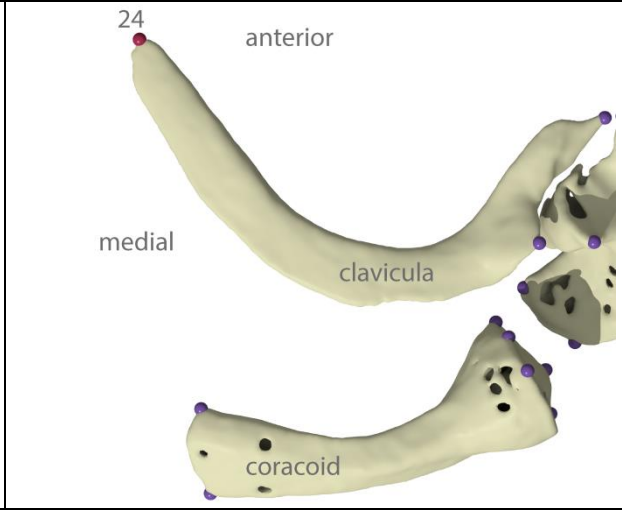
Algorithm	Radius	Parameter 1	Combination strategy	Misclassified voxels [%]
Bernsen	5	26	[One direction only]	4.2858
Bernsen	5	26	intersecting-2-of-3	4.0056
Bernsen	5	26	intersecting-3	3.6148
Bernsen	5	27	[One direction only]	4.2858
Bernsen	5	27	intersecting-2-of-3	4.0053
Bernsen	5	27	intersecting-3	3.6044
Bernsen	5	28	[One direction only]	4.2849
Bernsen	5	28	intersecting-2-of-3	4.0044
Bernsen	5	28	intersecting-3	3.5990
Bernsen	7	21	[One direction only]	4.3485
Bernsen	7	21	intersecting-2-of-3	4.0912
Bernsen	7	21	intersecting-3	3.4442
Bernsen	7	22	[One direction only]	4.3480
Bernsen	7	22	intersecting-2-of-3	4.0896
Bernsen	7	22	intersecting-3	3.4433
Bernsen	7	23	[One direction only]	4.3473
Bernsen	7	23	intersecting-2-of-3	4.0896
Bernsen	7	23	intersecting-3	3.4422
MidGrey	8	5	[One direction only]	7.8054
MidGrey	8	5	intersecting-2-of-3	6.8866
MidGrey	8	5	intersecting-3	5.6535
MidGrey	9	-4	[One direction only]	2.5170
MidGrey	9	-4	intersecting-2-of-3	2.5006
MidGrey	9	-4	intersecting-3	2.8699
MidGrey	9	5	[One direction only]	7.7562
MidGrey	9	5	intersecting-2-of-3	6.8900
MidGrey	9	5	intersecting-3	5.6454
MidGrey	10	-4	[One direction only]	2.5293
MidGrey	10	-4	intersecting-2-of-3	2.5663
MidGrey	10	-4	intersecting-3	2.7871
Otsu	5	0	[One direction only]	6.3128
Otsu	5	0	intersecting-2-of-3	5.3678
Otsu	5	0	intersecting-3	5.5612
Otsu	6	0	[One direction only]	6.3983
Otsu	6	0	intersecting-2-of-3	5.5452
Otsu	6	0	intersecting-3	5.0281
Otsu	14	0	[One direction only]	7.9991
Otsu	14	0	intersecting-2-of-3	7.3603
Otsu	14	0	intersecting-3	5.6204
Otsu	15	0	[One direction only]	8.0764
Otsu	15	0	intersecting-2-of-3	7.4760
Otsu	15	0	intersecting-3	5.7493

Supporting Information Table S2 Landmark definitions and visualizations.

Landmark	Description/remark	Visualization
1	Cleithrum, dorsal point of anterior margin. Set in frontomedial view.	
2	Cleithrum, ventral point of anterior margin. Set in frontal view.	
3	Scapula, most anterior point of dorsal face. A potential bony crest anterior to this point is neglected. Set in dorsal view.	
4	Scapula, most posterior point of dorsal face. Set in dorsal view.	

Landmark	Description/remark	Visualization
5	Scapula, most anterior point of posterior margin (extremal point of concave posterior margin). Set in lateral view.	
6	Scapula, point at the angulation between dorsal glenoid cavity margin and posterior margin of pars glenoidalis. Set in posteromedial view.	
7	Scapula, posteroventral point of pars glenoidalis at the angulation between posterior and ventral margin. Set in posteromedial view.	
8	Coracoid, posterodorsal point of glenoidal face. Set in posteromedial view.	
9	Scapula, anteroventral point of pars glenoidalis at the transition of ventral to anterior margin. Set in anteromedial view.	
10	Coracoid, anterodorsal point of glenoidal face. Usually there is a depression in the anterior margin extending between this point and landmark 19. Set in anteromedial view.	
11	Scapula, dorsal point of medial notch between partes acromialis and glenoidalis. Set in anteromedial view.	
12	Scapula, ventral point of posterior margin of medial surface of pars acromialis. Set in anteromedial view.	
13	Clavicula, lateral point of posterodorsal margin at the angulation to laterodorsal margin. Set in medial view.	
14	Scapula, ventral point of anterior margin. Set in anteromedial view.	

Landmark	Description/remark	Visualization
15	Clavicula, lateral point of anterior margin. Set in anterolateral view.	
16	Scapula, ventral point of posterior margin of lateral surface of pars acromialis. Set in anterolateral view.	
17	Clavicula, lateral point of posteroventral margin at angulation to lateroventral margin. Set in anterolateral view.	
18	Scapula, dorsal point of lateral incision between pars acromialis and pars glenoidalis. Set in lateral view.	
19	Coracoid, anterior margin at the transition of glenoidal to scapula face. There usually is a depression of the anterior coracoid margin extending between this point and landmark 10. Set in lateroventral view.	
20	Coracoid, anteroventral point of lateral/glenoidal surface of coracoid. Set in lateroventral view.	
21	Coracoid, posterior margin of glenoidal face at the projection of the coracoid long axis. Set on posterior view.	
22	Coracoid, posteromedial apex at angulation between medioventral and mediodorsal margin. Set in medial view.	
23	Coracoid, anteromedial apex at angulation between medioventral and mediodorsal margin. Set in medial view.	

Landmark	Description/remark	Visualization
24	Clavicula, anteromedial tip. Set in ventral view.	

Supporting Information Table S3 Overview of the composition of Landmark Datasets 1–4.

Landmark Dataset	1	2	3	4
Specimen(s) and girdle half/halves	9 *Bombina bombina*, 9 *Bombina orientalis*, both girdle halves	*Bombina orientalis* ZMH A12601, ventral girdle half	*Bombina orientalis* ZMH A12601, ventral girdle half	*Bombina orientalis* ZMH A05682, ZMH A05676, ventral girdle halves
Surface variant(s)	- Downsampling: NoDown - Segmentation: SubThresh, OtsuT, MidGreyT - Simplification: original, subSimpl	- Downsampling: NoDown, Down2, Down4 - Segmentation: SubThresh, OtsuT, MidGreyT, minThresh, maxThresh - Simplification: original, subSimpl, strongSimpl	One surface: - Downsampling: NoDown - Segmentation: MidGreyT - Simplification: original	- Downsampling: NoDown - Segmentation: SubThresh, OtsuT, MidGreyT - Simplification: original, subSimpl
Observers	O1, O2	O1	O1–O6	O1
Repetitions	3 repetitions on 3 different days (1 repetition per day)	9 repetitions on 3 different days (3 repetition per day)	30 repetitions on 2 consecutive days (15 repetitions per day)	9 repetitions in 3 different sessions (3 repetitions per session)

Supporting Information Text S1: Comparison of automatic local thresholding strategies

To compare the performance of different combinations of automatic local thresholding algorithms and respective parameter, a phantom stack was created based on a 480x480x286 voxel sub-volume of the scan of specimen ZMH A12601. The sub-volume contained the pectoral girdle bones and parts of the surrounding calcified and soft tissues and served to assign each pixel in the phantom images to its respective structure. Each phantom image pixel belonging to a left side pectoral girdle bone was given the gray value that equaled the mean gray value of the respective bone in the original µCT scan. All other bone-pixels in the phantom stack were given an uniform gray value. Soft tissue and ethanol drops in the wadding were represented by the mean gray value of the respective regions in the original scan. The phantom stack was created in Amira® and exported as an image stack of transversal slices.

Each phantom image was then Radon-transformed to obtain a sinogram (700 x 481 pixels), which in turn was inversely Radon-transformed (functions *radon* and *iradon* with default settings but adjusted pixel counts; *PET* package [version 0.5.0; Schulz et al. 2018] for R [version 3.4.3; R Core Team, 2017] using RStudio [version 1.1.383; RStudio Team, 2017]) to simulate a CT scan with 700 projections. The resulting stack of reconstructed phantom images was imported to Fiji; noise and blurring were added to obtain a more realistic stack ("reconstructed phantom stack/images"). The nine automatic thresholding algorithms implemented in the *Auto Local Threshold* function were applied to each reconstructed phantom image with systematically varying radius and parameters using a custom macro (see plugin documentation for details on algorithms and parameters). About 400 combinations of algorithms and parameters were tested. During the thresholding trials each pixel was classified either as object (white; here bone) or background (black).

The quality of the thresholding was assessed for the left-side pectoral girdle bones using two different measures adopted from Yasnoff et al. (1977). The first quality measure was the percentage of misclassified voxels of the bone volumes as derived from the original phantom stack and of the two rows of voxels adjacent to the bones (bone volume grown by two voxels in all three dimensions). The second quality measure was the percentage of misclassified voxels in the bone volume only. Percentages were calculated from counts of the misclassified relative to all voxels in the evaluated volume (determined in Fiji). Based on the results we selected the three most promising thresholding algorithms (*MidGrey*, *Bernsen*, *Otsu*) with the parameter combinations that resulted in the best and second-best values (Supporting Information Table S1) for each of them and each of the quality measures. For the *MidGrey*

algorithm, the second measure (bone only) decreased without reaching a minimum while the other measure reached unreasonably high values. In this case, we considered only those parameter combinations that resulted in quality values that were comparable to the results of the other algorithms.

To account for the three-dimensionality of the data, the reconstructed phantom stack was resliced from both left to right and top to bottom in Fiji. Each of the two resulting stacks was subject to the three best thresholding algorithms with their best and second-best parameter combinations. The thresholding results were resliced to their original orientation such that they showed cross-sections again. The three corresponding thresholded stacks of each algorithm-parameter-combination served to create two new volumes: the first volume contained white voxels (value 1) only for those voxels that were recognized as object in any two of the three thresholded stacks ("intersecting-2-of-3 strategy"). The second contained white voxels only for those voxels that were recognized as object in all three thresholded stacks ("intersecting-3 strategy"). The combination of thresholded stacks was done in Amira® using the *Arithmetic* module with the expressions "((*a* > 0) && (*b* > 0)) || ((*a* > 0) && (*c* > 0)) || ((*b* > 0) && (*c* > 0))" and "(*a* > 0) && (*b* > 0) && (*c* > 0)", respectively, with *a*, *b*, and *c* being the (re-resliced) thresholding results. Thresholding quality was assessed by the percentage of misclassified voxels in the bone volume and in the volume of the two rows of voxels adjacent to the bones of the left-side pectoral girdle (first measure described above), but raw voxel counts were determined in Amira®.

The *MidGrey* algorithm with a radius of 9, the parameter -4, and the intersecting-2-of-3 strategy performed best with a misclassification rate of 2.5 % in the evaluated volume (Supporting Information Table S1).

References cited in the Supporting Information

R Core Team (2017) *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

RStudio Team (2017) *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. Available from: http://www.rstudio.com/.

Schulz J, Toft P, Jensen JJ, Philipsen P (2018) *PET: simulation and reconstruction of PET images*. Available from: https://CRAN.R-project.org/package=PET.

Yasnoff WA, Mui JK, Bacus JW (1977) Error measures for scene segmentation. *Pattern Recognit* **9**, 217–231.

Supporting Information Chapter three

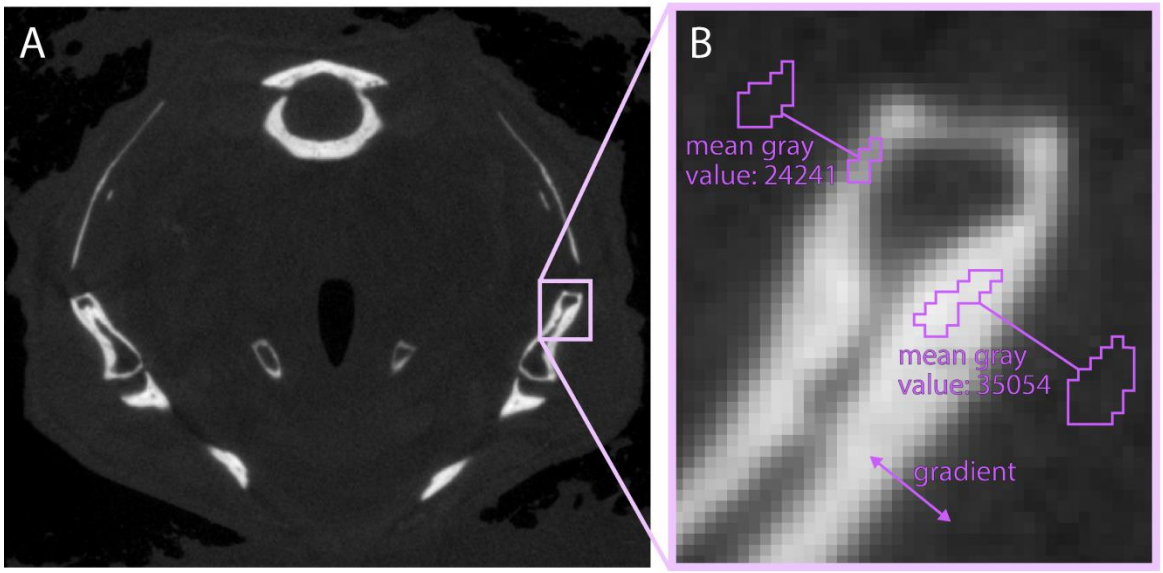


Fig. S1 Cross section of µCT volume of *Bombina orientalis* (ZMH A12601) illustrating the characteristics of such data. (A) Overview. (B) Detail of (A) showing a gray value gradient between bone and soft tissue, and locally differing mean gray values of bone and soft tissue.

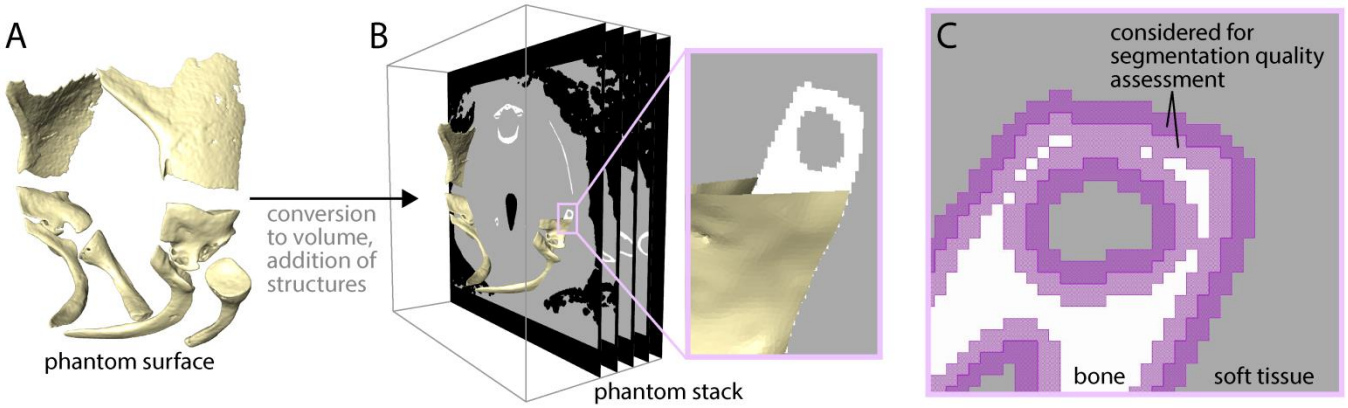


Fig. S2 Phantom surface and phantom stack. (A) Phantom surface used to create the phantom stack. (B) Phantom stack superimposed with phantom surface (cut) and detail showing a well-defined boundary between bone and soft tissue. (C) Detail of phantom stack with voxels (violet) considered for assessing the segmentation quality.

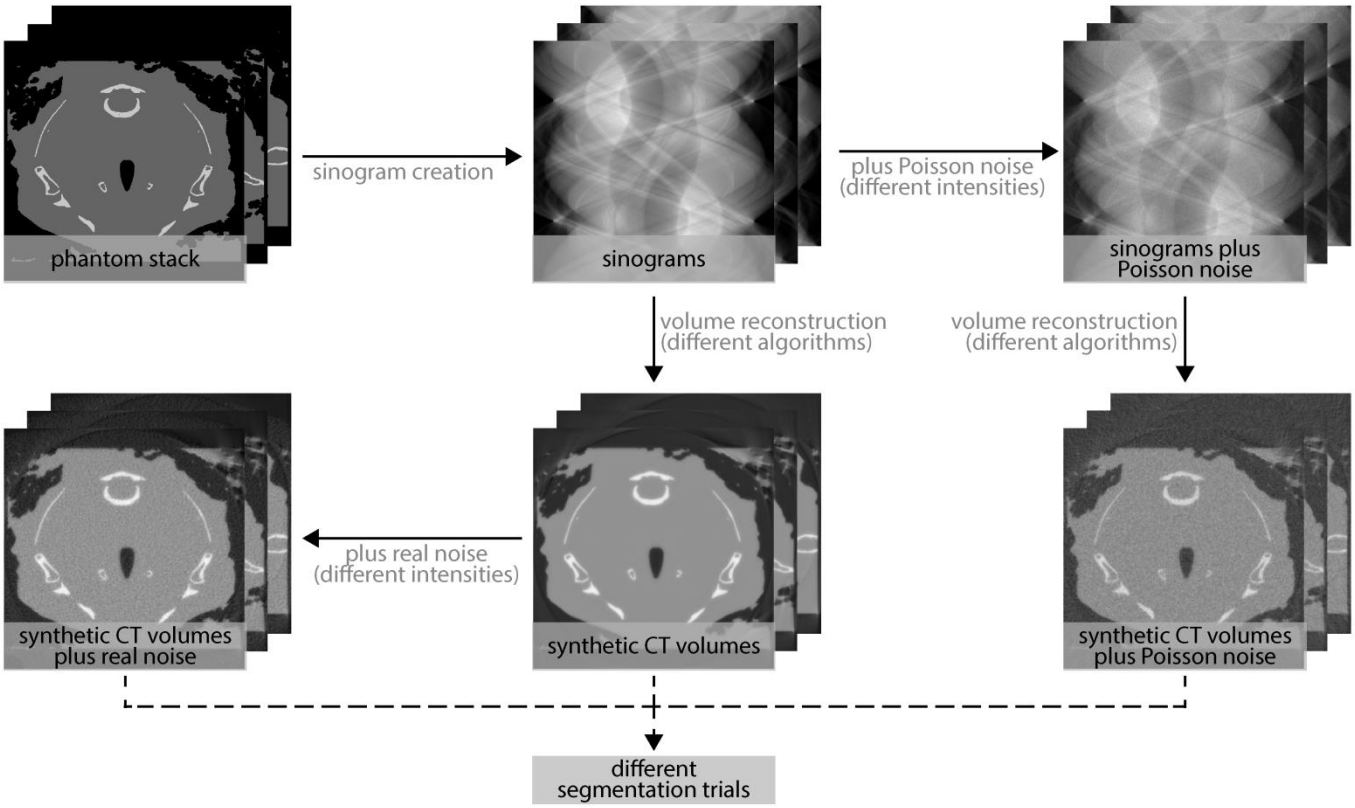


Fig. S3 Workflow to generate synthetic CT volumes. The synthetic volumes were used in the segmentation trials.

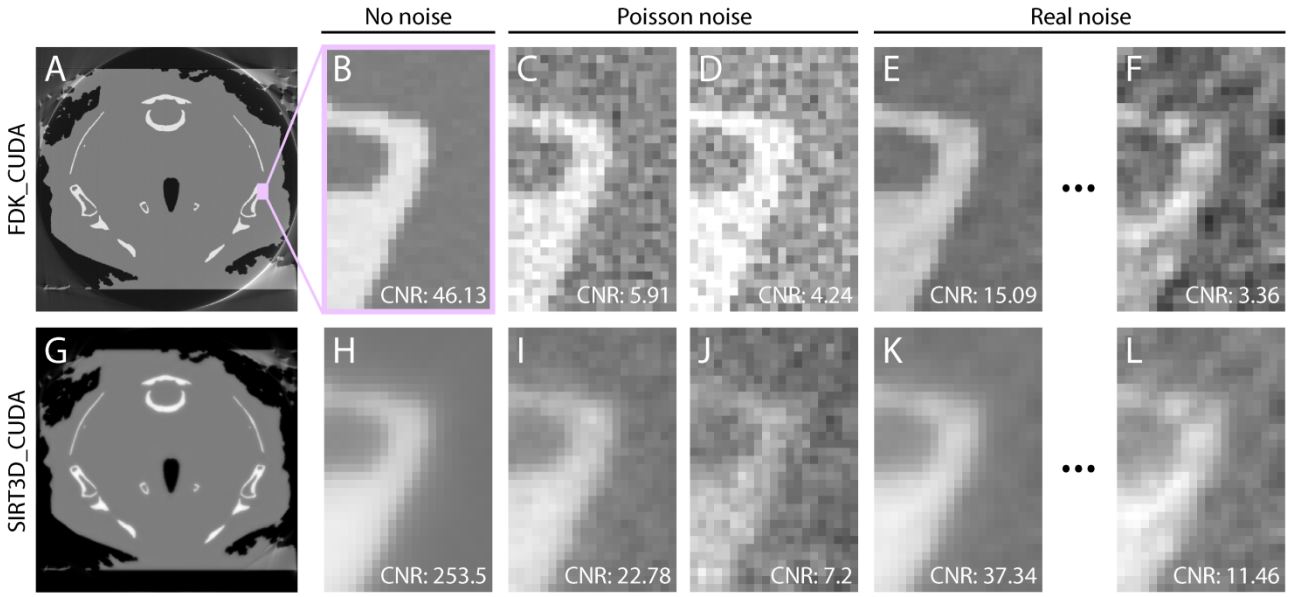


Fig. S4 Slices of selected synthetic CT volumes showing different lengths of the gray value gradient between structures, gray value distributions within bones, and noise characteristics. No noise: no noise added. Poisson noise: Poisson noise added to sinograms before volume reconstruction. Real noise: noise of real CT volume added to corresponding volume without noise. CNR: contrast to noise ratio calculated for bone and soft tissue. (A) Slice of synthetic volume that was reconstructed using the algorithm *FDK_CUDA*. (B)-(F) Details of synthetic volumes with same reconstruction algorithm as (A) but different noises added. (G) Slice of synthetic volume that was reconstructed using the algorithm *SIRT3D_CUDA*. (H)-(L) Details of synthetic volumes with same reconstruction algorithm as (G) but different noises added.

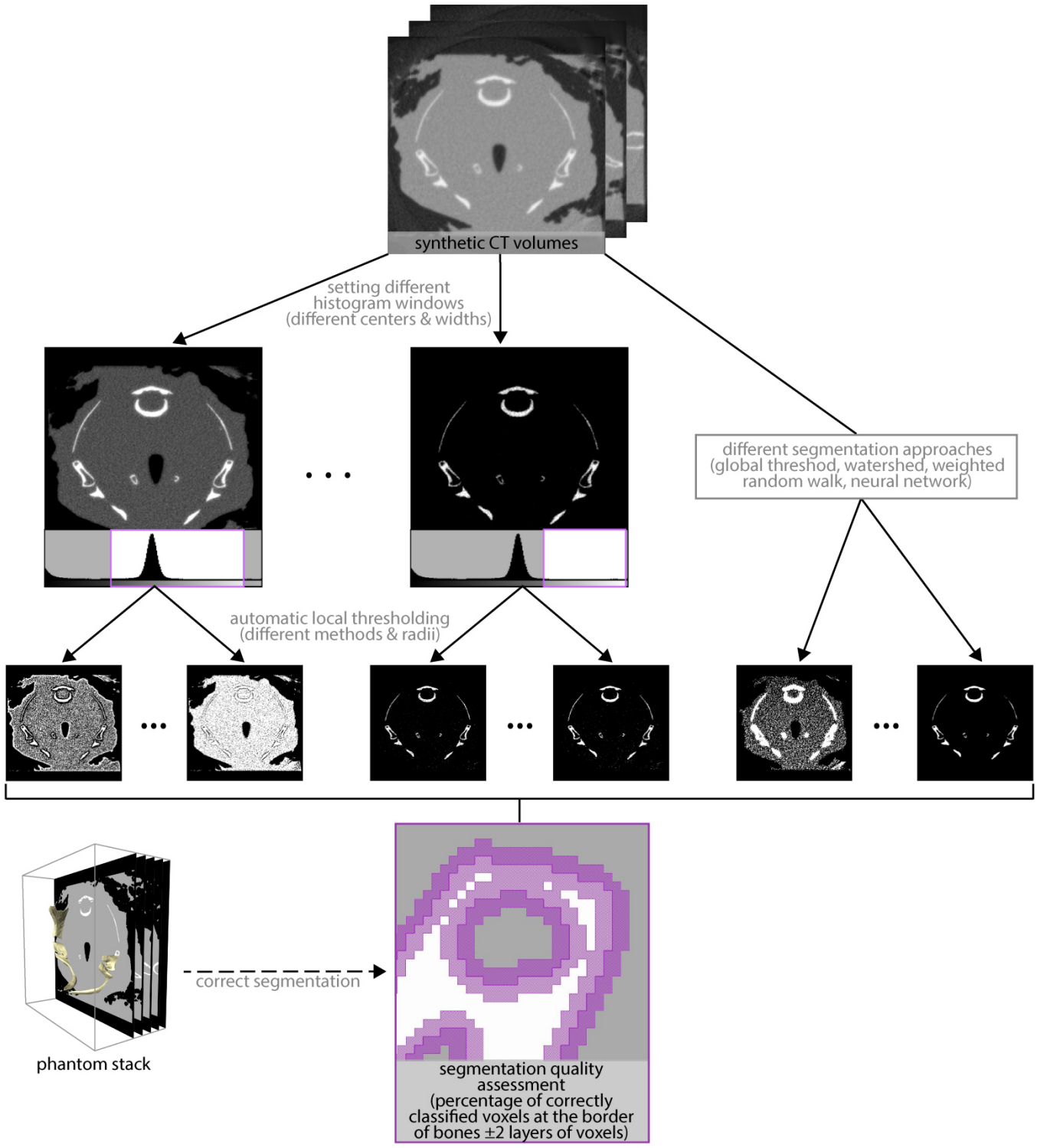


Fig. S5 Workflow of performing segmentation trials to assess the performance (quality) of different approaches.

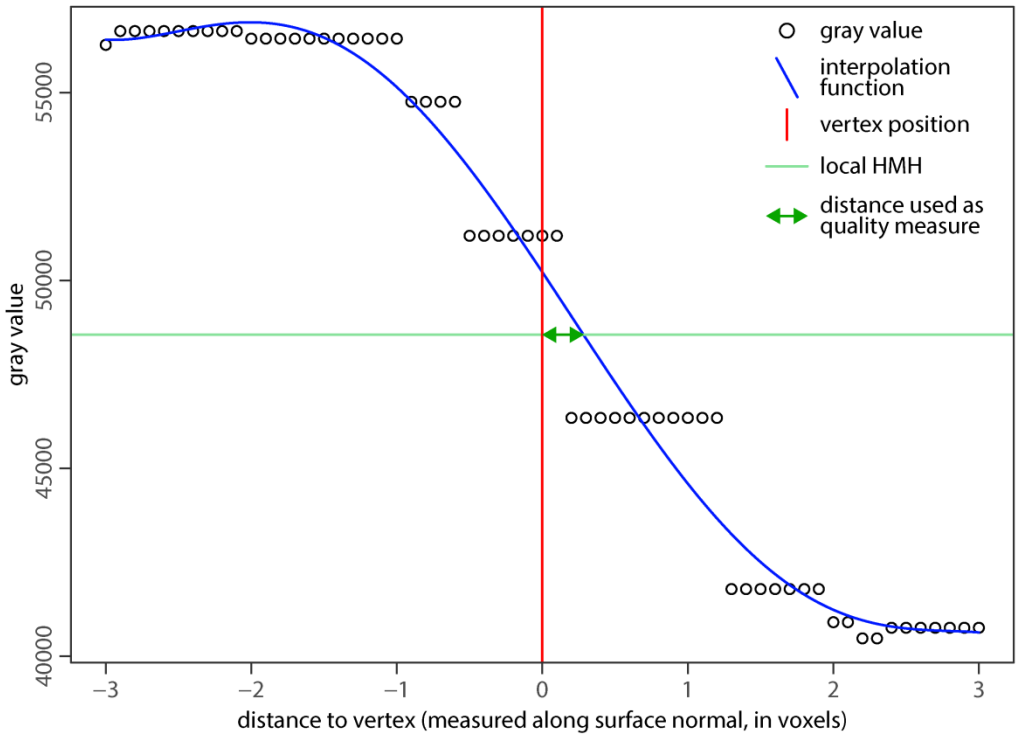


Fig. S6 Illustration of the accuracy assessment of a surface derived from real µCT volumes. The gray values (circles) were extracted along the surface normal (*x*-axis) of a randomly selected vertex (red line). The extracted gray values were interpolated by a B-spline (degree = 5; blue curve) and the local half maximum height (HMH, light green line) was determined. The intersection of the blue curve and the light green line marks the location of the true boundary between bone and soft tissue. The distance (green arrow) between the location of the true boundary and the position of the vertex was considered as approximation for the accuracy of the surface; small distances indicate a high surface accuracy.

Fig. S7 (part 1) Segmentation and surface accuracies of pectoral girdle bones in real µCT scans. Details of CT volumes show the segmentation results (violet) of the different segmentation approaches. Surfaces were generated with *Generate Lego Surface* module and simplified to a subjectively optimal degree in Amira® (approach i). Boxplots illustrate the estimated accuracy of the simplified surfaces (expressed as the distance of selected vertices to the location of the half maximum height (HMH)). Note that this estimation is prone to different kinds of errors. One unit equals one voxel. Potential outliers not plotted.

Fig. S7 (part 2) Segmentation and surface accuracies of pectoral girdle bones in real µCT scans. Details of CT volumes show the segmentation results (violet) of the different segmentation approaches. Surfaces were generated with *Generate Lego Surface* module and simplified to a subjectively optimal degree in Amira® (approach i). Boxplots illustrate the estimated accuracy of the simplified surfaces (expressed as the distance of selected vertices to the location of the half maximum height (HMH)). Note that this estimation is prone to different kinds of errors. One unit equals one voxel. Potential outliers not plotted.

Fig. S7 (part 3) Segmentation and surface accuracies of pectoral girdle bones in real µCT scans. Details of CT volumes show the segmentation results (violet) of the different segmentation approaches. Surfaces were generated with *Generate Lego Surface* module and simplified to a subjectively optimal degree in Amira® (approach i). Boxplots illustrate the estimated accuracy of the simplified surfaces (expressed as the distance of selected vertices to the location of the half maximum height (HMH)). Note that this estimation is prone to different kinds of errors. One unit equals one voxel. Potential outliers not plotted.

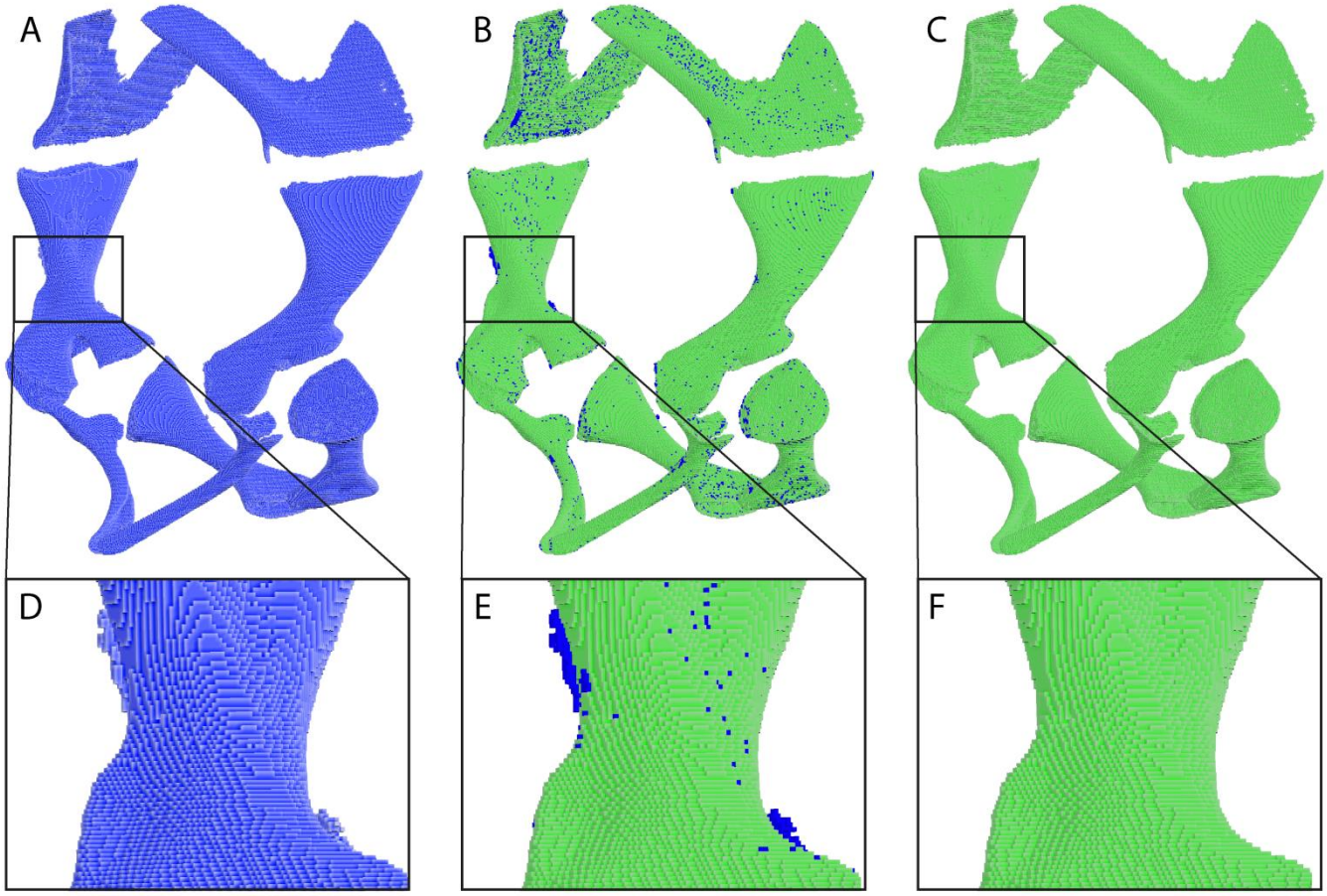


Fig. S8 Comparison of unsimplified surfaces of *Ecnomiohyla miliaria* derived from different segmentation results obtained by automatic local thresholding. Thresholding parameters were chosen as recommended in the text. Surfaces generated by *Generate Lego Surface* module in Amira®. (A) Thresholding performed on cross section images (cross). (B) Surface of (C) visualized in green on top of surface of (A) in blue to highlight inaccuracies (noise) in surface areas parallel to the plane of thresholding, if thresholding is performed in one plane only (cross). (C) Thresholding performed in three different planes and thresholding results combined by setting those voxels as bone voxels that were recognized as bone in any two of the three thresholding results (2of3). (D)-(E) Details of (A)-(C).

Table S1 CT volumes downloaded from MorphoSource and used in study.

Species	Institution	Catalogue number	MorphoSource media number
Ecnomiohyla miliaria	University of Florida, Florida Museum of Natural History	UF-Herpetology 137208	M25112-49249
Mixophyes fasciolatus fasciolatus	California Academy of Sciences	CAS 82050	M23916-47036
Occidozyga baluensis	University of Kansas Natural History Museum	KUH 155619	M25762-50252

Table S2 Highest achieved segmentation quality of automatic local and global thresholding and respective parameters used to obtain best thresholding results.

Characteristics of synthetic CT volumes[a]								Automatic local thresholding[b]					
								Otsu				2of3[c]	3of3[d]
Reco	**Noise**	**Mean** soft tissue	**SD** soft tissue	**Mean** bone	**SD** bone	**Mean** bone & soft issue	**CNR**	**Quality [%]**	**Center**	**Width**	**r**	Quality [%]	Quality [%]
FDK	No	16726	222.1	26972	244.1	21849	46.1	99.9928	21750	5300	4	99.9866	99.9883
FDK	Real	16813	683.4	27123	695.4	21968	15.1	99.8676	21950	4400	3	(not tested)	
FDK	Real	16892	1119.1	27253	1136.4	22072	9.3	99.1149	22050	4500	3	99.1502	99.0467
FDK	Poisson	17691	1601.0	27149	1634.8	22420	5.9	97.1811	22450	1900	18	(not tested)	
FDK	Real	17014	2166.6	27456	2247.6	22235	4.8	93.9531	22250	1000	19	93.8515	93.8506
FDK	Poisson	19897	2096.8	28786	2001.4	24342	4.2	93.4427	24450	1400	28	93.4433	93.4400
FDK	Real	17158	3240.6	28054	3267.3	22606	3.4	87.0069	22400	1200	28	(not tested)	
SIRT	No	37348	94.5	61302	2397.4	49325	253.5	99.2343	49950	15100	3	99.0304	97.0666
SIRT	Real	37431	645.8	61546	2133.6	49489	37.3	98.0971	49900	13900	3	(not tested)	
SIRT	Poisson	37566	999.7	60340	2159.9	48953	22.8	96.3169	49650	13800	5	96.0013	94.6419
SIRT	Real	37492	1068.5	61750	2293.0	49621	22.7	96.3868	50150	14100	4	96.1302	94.0895
SIRT	Real	37546	1511.6	62307	2018.6	49926	16.4	94.2811	50200	14000	5	(not tested)	
SIRT	Real	37633	2157.4	62355	2382.1	49994	11.5	90.8477	50400	14200	7	90.6593	89.6601
SIRT	Poisson	36421	2533.7	54671	3042.6	45546	7.2	84.5155	46050	10800	9	(not tested)	

[a] **Reco**: reconstruction algorithm used during generation of synthetic CT volumes (FDK: FDK_CUDA; SIRT: SIRT3D_CUDA); **Noise**: Type of noise added to the synthetic CT volumes; No: no noise added; Poisson: Poisson noise added to sinograms; Real: noise extracted from real CT volume and added to synthetic stacks without noise; **Mean/SD** soft tissue: mean/standard deviation of gray values of soft tissue; **Mean/SD** bone: mean/standard deviation of gray values of bone; **Mean** bone & soft tissue: mean gray value of bone and soft tissue; **CNR**: contrast to noise ratio calculated for bone and soft tissue

[b] Automatic local thresholding performed using the different methods (*Otsu*, ...); **Quality [%]**: segmentation quality measured as percentage of correctly classified voxels at the boundary of the bones ±2 rows of voxels; **Center/Width**: parameters determining the center/width of the histogram window set prior to automatic local thresholding; **r**: radius set for automatic local thresholding

[c] Resliced versions of volume thresholded (*Otsu* method with respectively optimal histogram window and radius) and combined by setting those voxels as bone, that were recognized as bone voxels in any two of the three thresholded stack versions

[d] Resliced versions of volume thresholded (*Otsu* method with respectively optimal histogram window and radius) and combined by setting those voxels as bone, that were recognized as bone voxels in all three thresholded stack versions

Table S2 continued.

Automatic local thresholding[b]											
Bernsen				*Mean*				*MidGrey*			
Quality [%]	**Center**	**Width**	**r**	**Quality [%]**	**Center**	**Width**	**r**	**Quality [%]**	**Center**	**Width**	**r**
99.9877	21750	4500	1	99.9908	22300	2600	11	99.9791	21650	5300	11
99.8477	21950	2700	1	99.8538	22250	1000	23	99.8435	21950	4100	10
99.0861	22050	2700	6	99.0952	22250	900	13	99.0844	22050	0	15
97.1802	22450	4300	29	97.1894	22700	700	22	97.1816	22450	300	14
93.9516	22250	1600	10	93.9601	22500	900	19	93.9486	22250	1100	16
93.4434	24450	4000	28	93.4419	24450	100	23	93.4433	24450	4100	27
87.0037	22500	2300	13	87.0046	22450	100	13	87.0040	22500	3800	9
99.5737	50400	16800	1	99.6582	53450	21900	1	99.4761	52100	20800	1
98.4423	50150	15200	1	98.2059	50750	14900	2	98.1610	52100	19400	1
96.0789	49500	17000	3	96.3425	49500	9000	6	95.9957	49600	17400	3
96.2566	50100	16500	2	96.3258	50350	9700	6	96.0392	50750	18300	2
93.9021	50050	18600	4	94.2922	50450	9400	7	93.8644	50100	19100	4
90.4000	50050	20000	5	90.9220	50800	9400	8	90.4038	49900	20000	5
84.0597	45550	18000	7	84.5025	46250	5900	11	84.0636	45450	18600	8

[b] Automatic local thresholding performed using the different methods (*Otsu*, ...); **Quality [%]**: segmentation quality measured as percentage of correctly classified voxels at the boundary of the bones ±2 rows of voxels; **Center/Width**: parameters determining the center/width of the histogram window set prior to automatic local thresholding; **r**: radius set for automatic local thresholding

Table S2 continued.

Automatic local thresholding[b]								Global threshold	
Niblack				*Phansalkar*					
Quality [%]	**Center**	**Width**	**r**	**Quality [%]**	**Center**	**Width**	**r**	**Quality [%]**	**Threshold**
99.9908	22100	2500	13	99.9986	22550	4900	1	99.9774	21650
99.8575	22100	1100	13	99.8681	22050	2600	3	99.8408	21950
99.0952	22150	700	13	99.0978	22150	1200	5	99.0842	22050
97.1885	22500	300	14	97.1814	22450	0	2	97.1816	22450
93.9630	22500	1000	23	93.9688	22950	2100	30	93.9484	22250
93.4434	24450	100	18	93.4383	24700	900	30	93.4397	24450
87.0085	22500	400	28	87.0123	22750	1100	29	87.0067	22400
98.9089	49050	9500	6	99.4784	54200	23300	3	94.1470	48100
97.8421	49150	9100	6	98.0183	54950	21400	5	93.3834	48200
96.2911	48650	8100	7	96.5230	52650	20000	4	92.4933	47600
96.2723	49350	8700	7	96.4246	55200	21700	6	92.2796	48200
94.2673	49700	8800	8	94.2973	54050	18400	7	90.8536	48350
90.9297	49950	8600	9	90.8439	52900	14800	8	88.3253	48500
84.5204	45900	5700	12	84.4987	48100	10800	12	83.4580	44650

[b] Automatic local thresholding performed using the different methods (*Otsu*, ...); **Quality [%]**: segmentation quality measured as percentage of correctly classified voxels at the boundary of the bones ±2 rows of voxels; **Center/Width**: parameters determining the center/width of the histogram window set prior to automatic local thresholding; **r**: radius set for automatic local thresholding

Table S3 Image-specific segmentation quality achieved with different approaches of watershed segmentation; quality measured as percentage of correctly classified voxels at the border of the bones ±2 rows of voxels.

Characteristics of synthetic CT volumes					Segmentation quality [%]											
Reco[a]	Noise[b]	SD[c]	Mean[d]	CNR[e]	Amira: *Sobel Filter*[f]			Fiji: *Gradient (3D)*[g]			Fiji: *Canny Edge*[h]			Amira: *Watershed*[i]		
					Ots-1[j]	Pha-1[k]	Pha-2[l]	Otsu-1	Pha-1	Pha-2	Otsu-1	Pha-1	Pha-2	Otsu-1	Pha-1	Pha-2
FDK	No	222.1	21849	46.1	95.3774	95.3807	95.3023	95.3786	95.3783	95.1782	92.5547	92.5579	92.2099	94.8753	94.8781	94.5780
FDK	Real	1119.1	22072	9.3	95.0564	95.1358	94.7480	95.1559	95.2379	94.1668	(not tested)			(not tested)		
FDK	Real	2166.6	22235	4.8	93.5421	94.2848	88.0044	93.7629	94.5294	86.7479	90.9706	91.5768	89.6089	93.0105	93.7216	88.9660
FDK	Poisson	2096.8	24342	4.2	94.4526	94.6460	93.9104	93.8595	94.1009	85.0511	(not tested)			(not tested)		
SIRT	No	94.5	49325	253.5	93.4466	93.6174	93.3786	93.9974	94.1766	94.1211	88.1065	88.3014	87.5519	91.9541	92.1628	91.6611
SIRT	Poisson	999.7	48953	22.8	92.7641	93.2551	92.9541	93.0890	93.6455	92.4222	(not tested)			(not tested)		
SIRT	Real	1068.5	49621	22.7	92.6107	93.0940	92.5173	93.1427	93.6834	92.9872	(not tested)			(not tested)		
SIRT	Real	2157.4	49994	11.5	90.4139	91.7665	86.2860	90.6572	92.1341	85.4079	86.6565	87.6557	84.8407	89.4051	90.6956	85.7569

[a] Reconstruction algorithm used during generation of synthetic CT volumes
[b] Type of noise added to the synthetic CT volumes; No: no noise added; Poisson: Poisson noise added to sinograms; Real: noise extracted from real CT volume and added to synthetic stacks without noise
[c] Standard deviation of soft tissue gray values
[d] Mean gray value of bone and soft tissue
[e] Contrast to noise ratio calculated for bone and soft tissue
[f] Gradient calculated by *Sobel Filter* (applied in 3D) in Amira® used for watershed segmentation
[g] Gradient calculated by *Gradient (3D)* function in Fiji
[h] Gradient calculated by *Canny Edge* function (applied with detection mode *Volumetric*, all other parameters kept at default) in Fiji
[i] Gradient calculated in the *Watershed Segmentation* module of Amira®
[j] Seed region for bone: Otsu-1-Seeds (pectoral girdle bone voxels in the best segmentation result obtained by auto local thresholding using the *Otsu* method shrinked by one row of voxels)
[k] Seed region for bone: Phantom-1-Seeds (pectoral girdle bone voxels in the phantom stack shrinked by one row of voxels)
[l] Seed region for bone: Phantom-2-Seeds (pectoral girdle bone voxels in the phantom stack shrinked by two rows of voxels)

Supproting information Chapter four [Appendix S1]

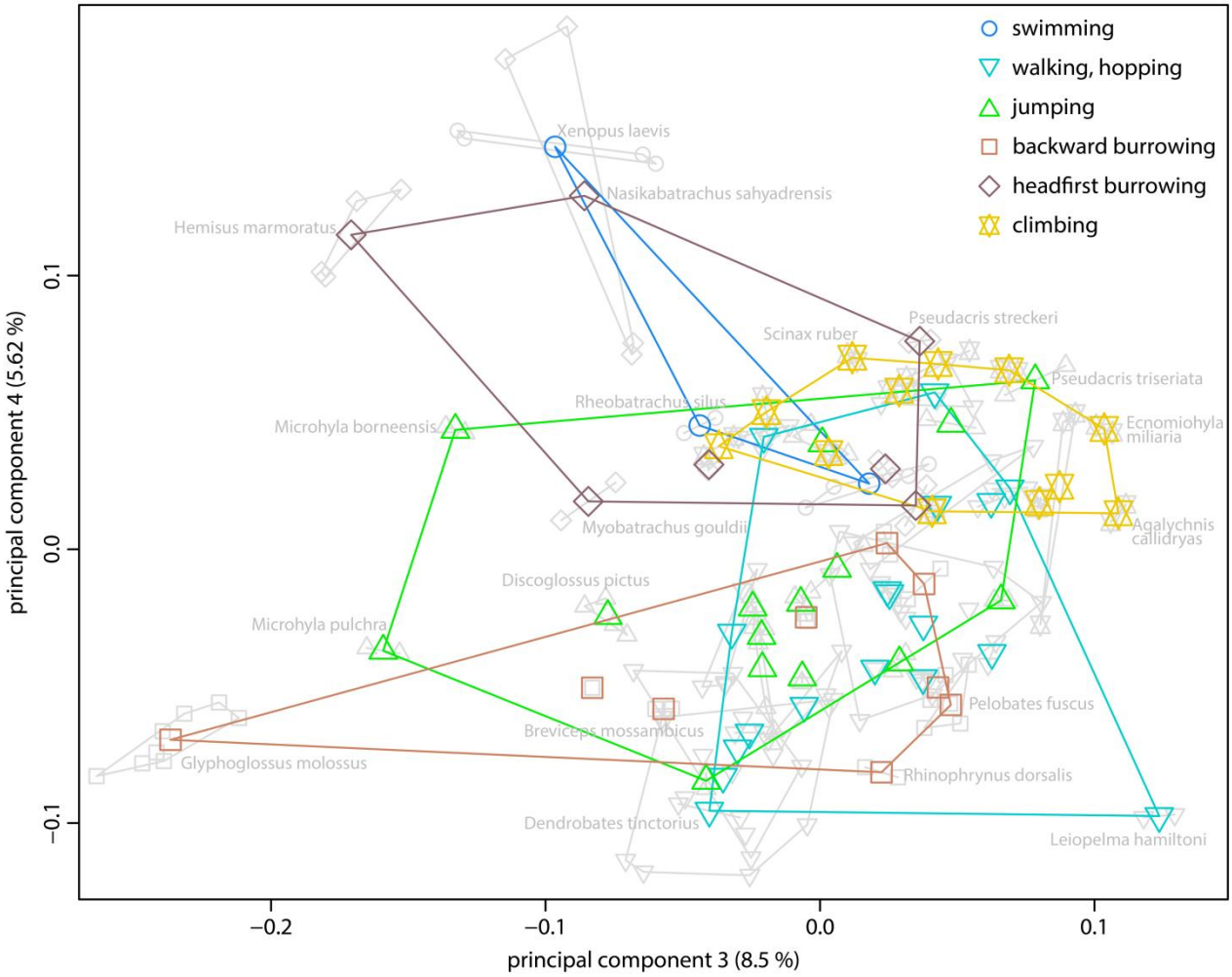


FIGURE A1 Plot of principal components 3 and 4 of overall species mean shapes of pectoral girdle bones (landmark dataset i). Grey points illustrate single landmark configurations used to calculate the species mean shapes.

TABLE A1 Specimens, locomotor modes, and CT scanning parameter. Museum abbreviations: AMNH: American Museum of Natural History, New York; CAS: California Academy of Sciences, San Francisco; Erfurt: Naturkundemuseum Erfurt, Erfurt; FMNH: Field Museum of Natural History, Chicago; USMN: National Museum of Natural History, Washington, D.C.; ZMB: Museum für Naturkunde, Berlin; ZMH: Zoologisches Museum Hamburg, Hamburg; ZSM: Zoologische Staatssammlung, Munich.

Species	**Locomotor mode**	**Reference**	**Catalogue number**	**CT scanning parameters [doi number]**	**Remark**
Afrixalus dorsalis (Peters, 1875)	climbing	Jorgensen & Reilly 2013	ZMB 71328	Skyscan1172; 52 kV; 188 µA; filter: Al 0.5 mm; voxel size: 13.33 µm [10.25592/uhhfdm.1142]	
Alytes obstetricans (Laurenti, 1768)	headfirst burrowing	Brown & Crespo 2000	ZMH A12442	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1144]	
Amietia angolensis (Bocage, 1866)	jumping	Laurent 1964; Emerson 1976	ZMH A07325	Skyscan1172; 55 kV; 179 µA; filter: Al 0.5 mm; voxel size: 22.40 µm [10.25592/uhhfdm.1146]	
Aplastodiscus leucopygius (Cruz and Peixoto, 1985)	climbing	Haddad & Sawaya 2000; Ferreira *et al.* 2008; Berneck, Segalla & Haddad 2017	USMN 208405	Skyscan1172; 52 kV; 188 µA; filter: Al 0.5 mm; voxel size: 21.08 µm [10.25592/uhhfdm.1148]	Headfirst burrowing in males for nest construction
Aromobates nocturnus Myers, Paolillo-O., and Daly, 1991	walking, hopping	Myers, Charles W., Paolillo O., Daly, John W. s W. & Daly 1991	AMNH A130017	Phoenix Nanotom S; 100 kV; 70 µA; no filter; voxel size: 30.00 µm [10.25592/uhhfdm.1152]	
Barbourula busuangensis Taylor and Noble, 1924	swimming	Myers 1943	CAS-SUA 21240	Skyscan1172; 100 kV; 100 µA; filter: Al+Cu; voxel size: 26.68 µm [10.25592/uhhfdm.1156]	
			CAS-SUA 21247	Skyscan1172; 100 kV; 100 µA; filter: Al+Cu; voxel size: 26.68 µm [10.25592/uhhfdm.1294]	
Bombina bombina (Linnaeus, 1761)	walking, hopping	Zug 1978; Cevik, Baskale & Kaya 2008	ZMH A05110	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.7455 µm [10.25592/uhhfdm.1162]	Scans from Engelkes *et al.* 2019
			ZMH A05383	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.75 µm [10.25592/uhhfdm.1164]	
			ZMH A05617	YXLON FF20 CT; 80 kV; 80 µA; no filter; voxel size: 25.84 µm [10.25592/uhhfdm.1166]	
			ZMH A09674	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1168]	Only left pectoral girdle half used; Scan from Engelkes *et al.* 2019
Bombina maxima (Boulenger, 1905)	walking, hopping	Jorgensen & Reilly 2013; Mai, Yu & Liao 2019	ZMH A05082	Skyscan1172; 70 kV; 139 µA; filter: Al+Cu; 26.68 µm [10.25592/uhhfdm.1170]	

TABLE A1 continued.

Species	Locomotor mode	Reference	Catalogue number	CT scanning parameters [doi number]	Remark
Bombina orientalis (Boulenger, 1890)	walking, hopping	Emerson 1979	ZMH A05672	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 30.33 µm [10.25592/uhhfdm.1172]	Scans from Engelkes *et al.* 2019
			ZMH A05676	Skyscan1172; 100 kV; 100 µA; filter: Al 0.5 mm; voxel size 21.34 µm [10.25592/uhhfdm.1174]	
			ZMH A05677	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.75 µm [10.25592/uhhfdm.1176]	
			ZMH A05678	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.75 µm [10.25592/uhhfdm.1182]	
			ZMH A05681	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.75 µm [10.25592/uhhfdm.1178]	Scans from Engelkes *et al.* 2019
			ZMH A05682	YXLON FF35 CT; 100 kV; 120 µA; no filter; voxel size: 22.75 µm [10.25592/uhhfdm.1180]	
Bombina variegata (Linnaeus, 1758)	walking, hopping	pers. obs. AH	ZMH A06234	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 17.60 µm [10.25592/uhhfdm.1184]	Only left pectoral girdle half used
Breviceps mossambicus Peters, 1854	backward burrowing	Poynton 1982	ZMB 83246	Skyscan1172; 55 kV; 181 µA; filter: Al 0.5 mm; voxel size: 16.54 µm [10.25592/uhhfdm.1186]	
Bufo bufo (Linnaeus, 1758)	walking, hopping	Zug 1978; Enriquez-Urzelai, Montori, Llorente & Kaliontzopoulou 2015	ZMH A04660	YXLON FF 35 CT; 70 kV; 120 µA; no filter; voxel size: 25.8 µm [10.25592/uhhfdm.1188]	
			ZMH A04680	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 63.41 µm [10.25592/uhhfdm.1190]	Only right pectoral girdle half used
			ZMH A04682	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 63.41 µm [10.25592/uhhfdm.1192]	
			ZMH A04708	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 63.41µm [10.25592/uhhfdm.1194]	
			ZMH A04709	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 63.41 µm [10.25592/uhhfdm.1198]	
			ZMH A04717	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 63.41 µm [10.25592/uhhfdm.1200]	
Ceratophrys aurita (Raddi, 1823)	backward burrowing	Nomura, Rossa-Feres & Langeani 2009; Natale *et al.* 2011	ZMH A01393	Phoenix v|tome|x L 450; 170 kV; 500 µA; no filter; voxel size: 106.01 µm [10.25592/uhhfdm.1204]	
Crossodactylus caramaschii Bastos and Pombal, 1995	headfirst burrowing	Nomura, Rossa-Feres & Langeani 2009	USMN 318234	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 11.87 µm [10.25592/uhhfdm.1206]	
Dendrobates tinctorius (Cuvier, 1797)	walking, hopping	Emerson 1979	ZMH A12904	Skyscan1172; 52 kV; 181 µA; filter: Al 0.5 mm; voxel size: 20.00 µm [10.25592/uhhfdm.1208]	

TABLE A1 continued.

Species	Locomotor mode	Reference	Catalogue number	CT scanning parameters [doi number]	Remark
Discoglossus montalentii Lanza, Nascetti, Capula, and Bullini, 1984	jumping	pers. obs. AH	ZSM 1299/2006	Skyscan1172; 52 kV; 181 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1212]	
			ZSM 1300/2006	Skyscan1172; 60 kV; 165 µA; filter: Al+Cu; voxel size: 21.08 µm [10.25592/uhhfdm.1214]	
Discoglossus pictus Otth, 1837	jumping	Emerson 1979	ZSM 933/2010	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 23.20 µm [10.25592/uhhfdm.1216]	
			ZSM 937/2010	Skyscan1172; 55 kV; 179 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1218]	
Discoglossus scovazzi Camerano, 1878	jumping	[based on other *Discoglossus* species]	ZMH A15451	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1220]	
Epidalea calamita (Laurenti, 1768)	backward burrowing	Emerson 1976, 1979	ZMH A06868	Skyscan1172; 100 kV; 100 µA; filter: Al 0.5 mm; voxel size: 22.94 µm [10.25592/uhhfdm.1292]	
Fejervarya limnocharis (Gravenhorst, 1829)	jumping	pers. obs. AH	ZMH A05523	Skyscan1172; 100 kV; 100 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1224]	
Gastrotheca riobambae (Fowler, 1913)	climbing	Hertwig & Sinsch 1995	CAS 119027	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1226]	
Glyphoglossus molossus Günther, 1869	backward burrowing	Emerson 1976	Erfurt A 1815/11	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 46.08 µm [10.25592/uhhfdm.1228]	
			Erfurt A 1818/11	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 46.08 µm [10.25592/uhhfdm.1230]	
			Erfurt A 1819/11	Phoenix Nanotom M; 120 kV; 400 µA; no filter; voxel size: 46.08 µm [10.25592/uhhfdm.1232]	
			Erfurt A 2186/15	Skyscan1172; 100 kV; 100 µA; filter: Al+Cu; voxel size: 26.68 µm [10.25592/uhhfdm.1234]	
Hemisus marmoratus (Peters, 1854)	headfirst burrowing	Emerson 1976	ZMB 79852 ZMH A06757	Skyscan1172; 55 kV; 181 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1236]	
				Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 18.14 µm [10.25592/uhhfdm.1238]	
Hyla arborea (Linnaeus, 1758)	climbing	pers. obs.; Cott 1926	ZMH A06468	Phoenix Nanotom S; 60 kV; 150 µA; no filter; voxel size: 23.23 µm [10.25592/uhhfdm.1240]	Only right pectoral girdle half used
Hyperolius parallelus Günther, 1858	climbing	[Channing 2001 for the genus *Hyperolius*]	ZMH A09562	Skyscan1172; 55 kV; 125 µA; filter: Al 0.5 mm; voxel size: 16.54 µm [10.25592/uhhfdm.1242]	
Kaloula pulchra Gray, 1831	backward burrowing	Emerson 1976	CAS 230419	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 18.68 µm [10.25592/uhhfdm.1244]	Only left pectoral girdle half used

TABLE A1 continued.

Species	Locomotor mode	Reference	Catalogue number	CT scanning parameters [doi number]	Remark
Kassina senegalensis (Duméril and Bibron, 1841)	walking, hopping	Emerson 1979	ZMB 75810	Skyscan1172; 52 kV; 188 µA; filter: Al 0.5 mm; voxel size: 19.74 µm [10.25592/uhhfdm.1246]	
			ZMH A07354	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 17.87 µm [10.25592/uhhfdm.1248]	
Leiopelma hochstetteri Fitzinger, 1861	walking, hopping	Worthy 1987	CAS-SUA 9609	Skyscan1172; 70 kV; 139 µA; filter: Al+Cu; voxel size: 20.27 µm [10.25592/uhhfdm.1250]	
Leptobrachella mjobergi Smith, 1925	walking, hopping	Hennigan 2013	ZMH A11518	Skyscan1172; 51 kV; 194 µA; filter: Al 0.5 mm; voxel size: 13.33 µm [10.25592/uhhfdm.1252]	
Leptodactylus pentadactylus (Laurenti, 1768	jumping	Emerson 1979	ZMH A02559	Phoenix v|tome|x L 450; 170 kV; 500 µA; no filter; voxel size: 91.40 µm [10.25592/uhhfdm.1254]	
Microhyla nepenthicola Das and Haas, 2010	jumping	pers. obs. AH	ZMH A11645	YXLON FF20 CT; 60 kV; 110 µA; no filter; voxel size: 12.69 µm [10.25592/uhhfdm.1256]	
Microhyla pulchra (Hallowell, 1861)	jumping	Emerson 1976	USMN 278542	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 13.33 µm [10.25592/uhhfdm.1258]	
Occidozyga baluensis (Boulenger, 1896)	walking, hopping	pers. obs. AH	ZMH A10454	Skyscan1172; 51 kV; 194 µA; filter: Al 0.5 mm; voxel size: 13.33 µm [10.25592/uhhfdm.1260]	
Oreobates quixensis Jiménez de la Espada, 1872	jumping	Jorgensen & Reilly 2013	AMNH A94687	Skyscan1172; 100 kV; 100 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1262]	
Pelobates fuscus (Laurenti, 1768)	backward burrowing	Savage 1942	ZMH A07151	Skyscan1172; 70 kV; 139 µA; filter: Al+Cu; voxel size: 26.68 µm [10.25592/uhhfdm.1266]	
Pelodytes punctatus (Daudin, 1802)	jumping	Enriquez-Urzelai, Montori, Llorente & Kaliontzopoulou 2015	ZMH A07281	Skyscan1172; 55 kV; 165 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1268]	
Pleurodema bibroni Tschudi, 1838	walking, hopping	Jorgensen & Reilly 2013	FMNH 132507	Skyscan1172; 55 kV; 181 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1270]	
Pseudacris streckeri Wright and Wright, 1933	headfirst burrowing	Brown, Jackson & Brown 1972	AMNH A184936	Skyscan1172; 55 kV; 179 µA; filter: Al 0.5 mm; voxel size: 21.34 µm [10.25592/uhhfdm.1272]	
Pseudacris triseriata (Wied-Neuwied, 1838)	jumping	Emerson 1979	CAS 188145	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 18.68 µm [10.25592/uhhfdm.1274]	
Rana temporaria Linnaeus, 1758	jumping	pers. obs.	ZMH A11310	Skyscan1172; 100 kV; 55 µA; no filter; voxel size: 26.18 µm [10.25592/uhhfdm.1278]	
Rentapia hosii (Boulenger, 1892)	climbing	pers. obs. AH	FMNH 244892	Skyscan1172; 54 kV; 185 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1264]	

TABLE A1 continued.

Species	Locomotor mode	Reference	Catalogue number	CT scanning parameters [doi number]	Remark
Rhacophorus nigropalmatus Boulenger, 1895	climbing	Emerson & Koehl 1990	ZMH A10414	YXLON FF35 CT; 60 kV; 160 µA; no filter; voxel size: 20.02 µm [10.25592/uhhfdm.1282]	Gliding
Rhinella marina (Linnaeus, 1758)	walking, hopping	Emerson 1979	ZMH A01033	Phoenix v|tome|x L 450; 170 kV; 500 µA; no filter; voxel size: 85.74 µm [10.25592/uhhfdm.1284]	
Rhinoderma darwinii Duméril and Bibron, 1841	jumping	Emerson 1979	ZMH A10873	Skyscan1172; 55 kV; 181 µA; filter: Al 0.5 mm; voxel size: 13.33 µm [10.25592/uhhfdm.1286]	
Rhinophrynus dorsalis Duméril and Bibron, 1841	backward burrowing	Trueb & Gans 1983	CAS 71767	Skyscan1172; 70 kV; 139 µA; filter: Al+Cu; voxel size: 21.60 µm [10.25592/uhhfdm.1288]	
Scinax ruber (Laurenti, 1768)	climbing	Pauly, Bernal & Taylor 2005	ZMH A02098	Skyscan1172; 51 kV; 192 µA; filter: Al 0.5 mm; voxel size: 21.08 µm [10.25592/uhhfdm.1290]	Parachuting
Xenopus laevis (Daudin, 1802)	swimming	Emerson 1979	ZMH A02374	Skyscan1172; 100 kV; 100 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [10.25592/uhhfdm.1296]	
Zhangixalus prominanus (Smith, 1924)	climbing	Shahrudin 2017	[one of the specimens in Barnes, Baum, Peisker & Gorb 2013]	Skyscan1172; 49 kV; 200 µA; filter: Al 0.5 mm; voxel size: 26.68 µm [unpublished]	

TABLE A2 Specimens, locomotor modes, and MorphoSource media number. Museum abbreviations: CAS: California Academy of Sciences, San Francisco; CES: Centre for Ecological Science, Bangalore; CM: Carnegie Museum of Natural History, Pittsburgh; Erfurt: Naturkundemuseum Erfurt, Erfurt; FMNH: Field Museum of Natural History, Chicago; KUH: University of Kansas Biodiversity Institute, Lawrence; MCZ: Museum of Comparative Zoology, Cambridge; UF: Florida Museum of Natural History, Gainesville.

Species	**Locomotor mode**	**Reference**	**Catalogue number**	**MorphoSource media number and URL**	**Remark**
Agalychnis callidryas (Cope, 1862)	climbing	Roberts 1994	CAS Herp 146957	M25577-50012, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25577	Parachuting
Alytes cisternasii Boscá, 1879	headfirst burrowing	Brown & Crespo 2000	MCZ A-3494	M35795-65967, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/35795	
Ascaphus truei Stejneger, 1899	jumping	Emerson 1979	UF Herp 80664	M8805-11256, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/8805, doi:10.17602/M2/M11256	
Bombina maxima (Boulenger, 1905)	walking, hopping	Jorgensen & Reilly 2013; Mai, Yu & Liao 2019	UF Herp 96648	M9207-23561, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/9207, doi:10.17602/M2/M23561	
Ecnomiohyla miliaria (Cope, 1886)	climbing	amphibiaweb.org (6th February 2020)	UF Herp 137208	M25112-49249, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25112	Gliding
Eleutherodactylus coqui Thomas, 1966	climbing	Stewart 1985	UF Herp 21290	M24647-48540, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/24647, doi:10.17602/M2/M48540	Parachuting
Epipedobates tricolor (Boulenger, 1899)	jumping	Jorgensen & Reilly 2013	KUH 219763	M24625-48513, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/24625	
			UF Herp 83888	M24980-49096, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/24980	
Eupsophus roseus (Duméril and Bibron, 1841)	walking, hopping	Meserve & Jaksic 1991	CM Herp 57175	M18659-35357, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/18659	
			CM Herp 63926	M12692-23439, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/12692, doi:10.17602/M2/M23439	
Gastrotheca riobambae (Fowler, 1913)	climbing	Hertwig & Sinsch 1995	UF Herp 98224	M28917-55509, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/28917	
Leiopelma hamiltoni McCulloch, 1919	walking, hopping	Worthy 1987	CAS Herp 53931	M13874-24351, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/13874, doi:10.17602/M2/M24351	
Leptobrachella mjobergi Smith, 1925	walking, hopping	Hennigan 2013	FMNH 273699	M23544-46055, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/23544	

TABLE A2 continued.

Species	Locomotor mode	Reference	Catalogue number	MorphoSource media number and URL	Remark
Leptobrachium hasseltii Tschudi, 1838	walking, hopping	Hennigan 2013	UF Herp 61841	M10832-16288, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/10832, doi:10.17602/M2/M16288	
Megophrys stejnegeri Taylor, 1920	walking, hopping	Hennigan 2013	KUH 321429	M22890-44409, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/22890	
Mixophyes fasciolatus Günther, 1864	walking, hopping	Littlejohn, Roberts, Watson & Davies 1993; Jorgensen & Reilly 2013	CAS Herp 82050	M23916-47036, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/23916	
Myobatrachus gouldii (Gray, 1841)	backward burrowing	Emerson 1976	MCZ A 139543	M25636-50088, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25636, doi:10.17602/M2/M50088	
Nasikabatrachus sahyadrensis Biju and Bossuyt, 2003	headfirst burrowing	Senevirathne *et al.* 2016	CES F 203	M12152-19736, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/12152	
			CES F 877	M12185-19838, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/12185	
Occidozyga baluensis (Boulenger, 1896)	walking, hopping	pers. obs. AH	KUH 155619	M25762-50252, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25762	
Pelobates fuscus (Laurenti, 1768)	backward burrowing	Savage 1942	UF Herp 36935	M25191-49377, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25191, doi:10.17602/M2/M49377	
Pyxicephalus adspersus Tschudi, 1838	backward burrowing	Loveridge & Withers 1981; pers. obs. AH	UF Herp 92094	M25376-49707, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25376	
Rheobatrachus silus Liem, 1973	swimming	Liem 1973; Littlejohn, Roberts, Watson & Davies 1993	CAS Herp 153753	M23917-47037, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/23917	
Sphaerotheca breviceps (Schneider, 1799)	backward burrowing	Nomura, Rossa-Feres & Langeani 2009	UF Herp 20069	M24970-49086, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/24970	
Xenopus laevis (Daudin, 1802)	swimming	Emerson 1979	CAS Herp 160540	M25472-49896, http://www.morphosource.org/Detail/MediaDetail/Show/media_id/25472, doi:10.17602/M2/M49896	

TABLE A3 Definition of landmarks and curves of semilandmarks (fixed landmarks adopted from Engelkes et al., 2019).

Number	Type	Definition/remark
L1	landmark	Cleithrum, dorsal point of anterior margin.
L2	landmark	Cleithrum, ventral point of anterior margin.
L3	landmark	Scapula, anterior point of dorsal margin.
L4	landmark	Scapula, posterior point of dorsal margin.
L5	landmark	Scapula, anterior point of posterior margin (extremal point of concave posterior margin).
L6	landmark	Scapula, pars glenoidalis, point on dorsal margin of glenoid cavity that is closest to L12.
L7	landmark	Scapula, pars glenoidalis, most posteroventral point of margin of glenoid cavity.
L8	landmark	Scapula, pars glenoidalis, anteroventral point at the transition of ventral to anterior margin.
L9	landmark	Scapula, dorsal point of medial notch between partes acromialis and glenoidalis.
L10	landmark	Scapula, pars acromialis, ventral point of posterior margin of medial surface.
L11	landmark	Scapula, ventral point of anterior margin.
L12	landmark	Scapula, pars acromialis, ventral point of posterior margin of lateral surface.
L13	landmark	Clavicula, lateralodorsal point of anterior margin.
L14	landmark	Clavicula, anteromedial point of anterior margin.
L15	landmark	Coracoid, lateral point of anterior margin.
L16	landmark	Coracoid, medial point of anterior margin.
L17	landmark	Coracoid, inflection point of anteroventral margin of glenoidal face.
L18	landmark	Coracoid, lateral point of posterior margin.
L19	landmark	Coracoid, lateral point of posterior margin.
C1	curve	Cleithrum, anterior margin, between L1 and L2; 29 semilandmarks plus endpoints.
C2	curve	Scapula, laterodorsal margin, between L3 and L4; 21 semilandmarks plus endpoints.
C3	curve	Scapula, posterior margin, between L4 and L5; 29 semilandmarks plus endpoints.
C4	curve	Scapula, margin of glenoid cavity, between L6 and L7; 25 semilandmarks plus endpoints.
C5	curve	Scapula, anterior margin, between L3 and L11; 29 semilandmarks plus endpoints.
C6	curve	Clavicula, anterior margin, between L13 and L14; 29 semilandmarks plus endpoints.
C7	curve	Coracoid, anterior margin, between L15 and L16; 29 semilandmarks plus endpoints.
C8	curve	Coracoid, ventral margin of glenoidal face, between L17 and L18; 26 semilandmarks plus endpoints.
C9	curve	Coracoid, anterior margin, between L18 and L19; 29 semilandmarks plus endpoints.

References cited in Appendix S1

Barnes, W.J.P., Baum, M., Peisker, H. & Gorb, S.N. (2013) Comparative Cryo-SEM and AFM studies of hylid and rhacophorid tree frog toe pads. *Journal of Morphology*, **274** (12), 1384–1396.

Berneck, B.v.M., Segalla, M.V. & Haddad, C.F.B. (2017) A first observation of amplexus in *Aplastodiscus* (Anura; Hylidae). *Herpetology Notes*, **10**, 351–354.

Brown, L.E. & Crespo, E.G. (2000) Burrowing behavior of the midwife toads *Alytes cisternasii* and *Alytes obstetricans* (Anura, Discoglossidae). *Alytes*, **17** (3-4), 101–113.

Brown, L.E., Jackson, H.O. & Brown, J.R. (1972) Burrowing behavior of the chorus frog, *Pseudacris streckeri*. *Herpetologica*, **28** (4), 325–328.

Cevik, I.E., Baskale, E. & Kaya, U. (2008) A mark-recapture study of the fire-bellied toad (*Bombina bombina*) using photographic recognition and dye marking techniques in North-wester Turkey. *Acta Biologica Universitatis Daugavpiliensis*, **8** (1), 75–79.

Channing, A. (2001) *Amphibians of Central and Southern Africa*. Comstock Publishing Associates, Ithaca.

Cott, H.B. (1926) Observations on the life-habits of some batrachians and reptiles from the Lower Amazon: and a note on some mammals from Marajó Island. *Proceedings of the Zoological Society of London*, **96** (4), 1159–1178.

Emerson, S.B. (1976) Burrowing in frogs. *Journal of Morphology*, **149** (4), 437–458.

Emerson, S.B. (1979) The ilio-sacral articulation in frogs: form and function. *Biological Journal of the Linnean Scociety*, **11**, 153–168.

Emerson, S.B. & Koehl, M.A.R. (1990) The interaction of behavioral and morphological change in the evolution of a novel locomotor type: "flying" frogs. *Evolution*, **44** (8), 1931–1946.

Engelkes, K., Helfsgott, J., Hammel, J.U., Büsse, S., Kleinteich, T., Beerlink, A., Gorb, S.N. & Haas, A. (2019) Measurement error in μCT-based three-dimensional geometric morphometrics introduced by surface generation and landmark data acquisition. *Journal of Anatomy*, **235** (2), 357–378.

Enriquez-Urzelai, U., Montori, A., Llorente, G.A. & Kaliontzopoulou, A. (2015) Locomotor mode and the evolution of the hindlimb in Western Mediterranean anurans. *Evolutionary Biology*, **42** (2), 199–209.

Ferreira, R.C., Souza, A.A. de, Freitas, R.A., Campaner, M., Takata, C.S.A., Barrett, T.V., Shaw, J.J. & Teixeira, M.M.G. (2008) A phylogenetic lineage of closely related trypanosomes (Trypanosomatidae, Kinetoplastida) of anurans and sand flies (Psychodidae, Diptera) sharing the same ecotopes in brazilian amazonia. *The Journal of eukaryotic microbiology*, **55** (5), 427–435.

Haddad, C.F.B. & Sawaya, R.J. (2000) Reproductive modes of Atlantic forest hylid frogs: a general overview and the description of a new mode. *BIOTROPICA*, **32** (4), 862.

Hennigan, T. (2013) An initial estimate toward identifying and numbering the frog kinds on the Ark: order Anura. *Answers Research Journal*, **6**, 335–365.

Hertwig, I. & Sinsch, U. (1995) Comparative toe pad morphology in marsupial frogs (genus *Gastrotheca*): arboreal versus ground-dwelling species. *Copeia*, **1995** (1), 38.

Jorgensen, M.E. & Reilly, S.M. (2013) Phylogenetic patterns of skeletal morphometrics and pelvic traits in relation to locomotor mode in frogs. *Journal of Evolutionary Biology*, **26** (5), 929–943.

Laurent, R.F. (1964) Adaptive modifications in frogs of an isolated highland fauna in Central Africa. *Evolution*, **18** (3), 458.

Liem, D.S. (1973) A new genus of frog of the family Leptodactylidae from SE Queensland, Australia. *Memoirs of the Queensland Museum*, **16**, 459–470.

Littlejohn, M.J., Roberts, J.D., Watson, G.F. & Davies, M. (1993) Family Myobatrachidae. *Fauna of Australia* (eds C.G. Galsby, G.J.B. Ross & P.L. Beesley). Australian government Publishing Service, Melbourne.

Loveridge, J.P. & Withers, P.C. (1981) Metabolism and water balance of active and Cocooned African Bullfrogs *Pyxicephalus adspersus*. *Physiological Zoology*, **54** (2), 203–214.

Mai, C., Yu, J. & Liao, W. (2019) Ecological and Geographical Reasons for the Variation of Digestive Tract Length in Anurans. *Asian Herpetological Research*, **10** (4), 246-252.

Meserve, P.L. & Jaksic, F.M. (1991) Comparisons of terrestrial vertebrate assemblages in temperate rainforests of North and South America. *Revista Chilena de Historia Natural*, **64**, 511–535.

Myers, G.S. (1943) Rediscovery of the Philippine discoglossid frog, *Barbourula busuangensis*. *Copeia*, **1943** (3), 148.

Myers, Charles W., Paolillo O., A., Daly, John W. s W. & Daly, J.W. (1991) Discovery of a malodorous and nocturnal frog in the family Dendrobatidae: phylogenetic significance of a new genus and species from the Venezuelan Andes. *American Museum Novitates*, **3002**, 1–20.

Natale, G.S., Alcalde, L., Herrera, R., Cajade, R., Schaefer, E.F., Marangoni, F. & Trudeau, V.L. (2011) Underwater acoustic communication in the macrophagic carnivorous larvae of *Ceratophrys ornata* (Anura: Ceratophryidae). *Acta Zoologica*, **92** (1), 46–53.

Nomura, F., Rossa-Feres, D.C. & Langeani, F. (2009) Burrowing behavior of *Dermatonotus muelleri* (Anura, Microhylidae) with reference to the origin of the burrowing behavior of Anura. *Journal of Ethology*, **27** (1), 195–201.

Pauly, G.B., Bernal, X. & Taylor, G.B. (2005) *Scinax ruber* (Red Snouted Treefrog). Arboreality and parachuting. *Herpetological Review*, **36** (3), 308–309.

Poynton, J.C. (1982) On species pairs among southern African amphibians. *African Zoology*, **17** (2), 67–74.

Roberts, W.E. (1994) Explosive breeding aggregations and parachuting in a neotropical frog, *Agalychnis saltator* (Hylidae). *Journal of Herpetology*, **28** (2), 193.

Savage, R.M. (1942) The burrowing and emergence of the Spade-Foot Toad, *Pelobates fuscus fuscus* Wagler. *Proceedings of the Zoological Society of London*, **A112** (1-2), 21–35.

Senevirathne, G., Thomas, A., Kerney, R., Hanken, J., Biju, S.D., Meegaskumbura, M. & Witten, P.E. (2016) From clinging to digging: the postembryonic skeletal ontogeny of the Indian Purple Frog, *Nasikabatrachus sahyadrensis* (Anura: Nasikabatrachidae). *PLoS ONE*, **11** (3), e0151114.

Shahrudin, S. (2017) Breeding activity of *Rhacophorus prominanus* Smith, 1924, from Peninsular Malaysia. *Herpetozoa*, **30** (1/2), 105–107.

Stewart, M.M. (1985) Arboreal habitat use and parachuting by a subtropical forest frog. *Journal of Herpetology*, **19** (3), 391–401.

Trueb, L. & Gans, C. (1983) Feeding specializations of the Mexican burrowing toad, *Rhinophrynus dorsalis* (Anura: Rhinophrynidae). *Journal of Zoology*, **199** (2), 189–208.

Worthy, T.H. (1987) Palaeoecological information concerning members of the frog genus *Leiopelma*: Leiopelmatidae in New Zealand. *Journal of the Royal Society of New Zealand*, **17** (4), 409–420.

Zug, G.R. (1978) Anuran locomotion-structure and function 2: jumping performance of semiaquatic, terrestrial and arboreal frogs. *Smithsonian Contributions to Zoology*, **276**, 1–31.

Supporting Information Chapter five [Electronic Supplementary Material]

Online Resource 1: Protocol for paraffin (Paraplast Plus®; Leica Biosystems) embedding of *Alytes obstetricans* (ZMH A12442). Steps adapted from Weninger et al. (1998); durations over 24 hours could have been reduced to 24 hours or less except for EDTA solution.

Time [h]	Step	Remarks
19	50 % ethanol	
9	30 % ethanol	
26	bidest. water	
	Specimen skinned	
75	Decalcification with ethylene-diamine-tetra-acetic acid solution (EDTA)	EDTA prepared following Mulisch and Welsch (2010) using ethylenediamine tetraacetic acid disodium salt dehydrate (Carl Roth GmbH + Co. KG)
57	EDTA changed	
20	EDTA changed	
4	Tab water	
18	Tab water changed	
45	30 % ethanol	
24	50 % ethanol	
24	70 % ethanol	
24	80 % ethanol + 2 g lead II acetate 3-hydrate/100 ml	Lead II acetate 3-hydrate obtained from Carl Roth GmbH + Co. KG
24	90 % ethanol + 2 g lead II acetate 3-hydrate/100 ml	
10	96 % ethanol + 2 g lead II acetate 3-hydrate/100 ml	
40	99.8 % ethanol	
46	Isopropanol	
31	Isopropanol changed	Kept in desiccator for 6 hours
24	Isopropanol : Paraplast Plus® 3:1	Three parts isopropanol mixed with one part Paraplast Plus®; kept at 60°C
21	Isopropanol : Paraplast Plus® 1:1	
27	Isopropanol : Paraplast Plus® 1:3	
22	Paraplast Plus®	
25	Paraplast Plus® changed	

References cited in Online Resource 1

Mulisch, Maria; Welsch, Ulrich (Eds.) (2010): Romeis Mikroskopische Technik. 18th ed. Heidelberg: Spektrum Akademischer Verlag.

Weninger, W. J.; Meng, S.; Streicher, J.; Müller, G. B. (1998): A new episcopic method for rapid 3-D reconstruction: applications in anatomy and embryology. In Anat Embryol 197 (5), pp. 341–348.

Online Resource 2:

Documentation of Amira® macro multiExport

The *multiExport* module iteratively calls the functions displayed in figure 1 to export versions of the original data set in a custom selection of the following data formats based on previously segmented *materials* of a *LabelField* (commonly also referred to as labels, labels object, or label data):

a. Separate image stacks (bmp, tif, or raw3d) containing either the original gray values or a "black-and-white-mask" of each selected structure/*material*.
b. Surfaces objects (obj or stl format) of each selected structure/*material*.
c. Simplified (reduced polygon count and/or smoothed) surfaces (obj or stl format) of each selected structure/*material*.

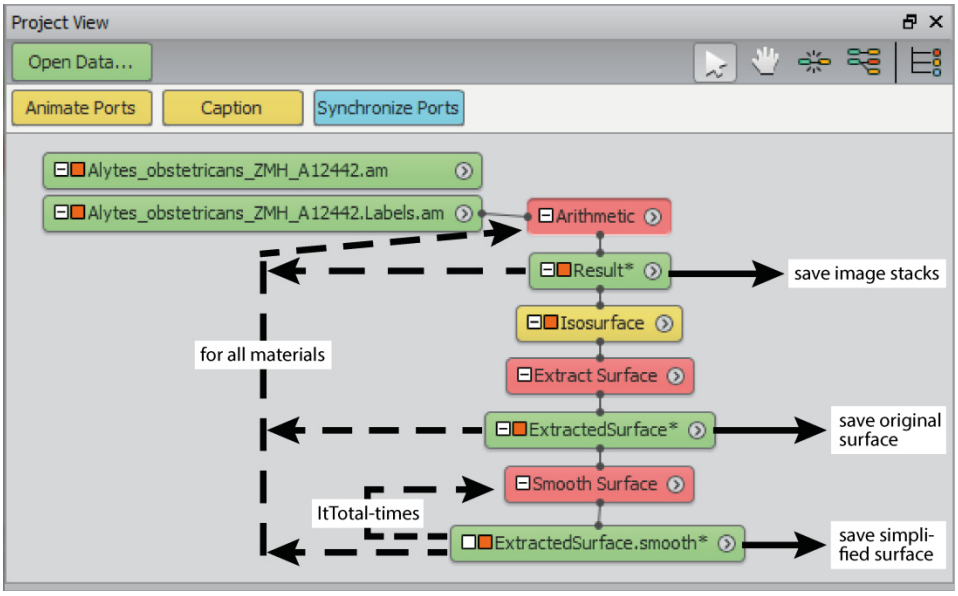


Figure 1: Maximal network created by *multiExport*. Some steps might be omitted depending on the output formats selected.

How to include *multiExport* in Amira®: Copy *multiExport.rc* and *multiExport.scro* in the *script-objects* folder of Amira® (usually found in C:/Programs/Amira-6.0.1/share/script-objects) and restart Amira®.

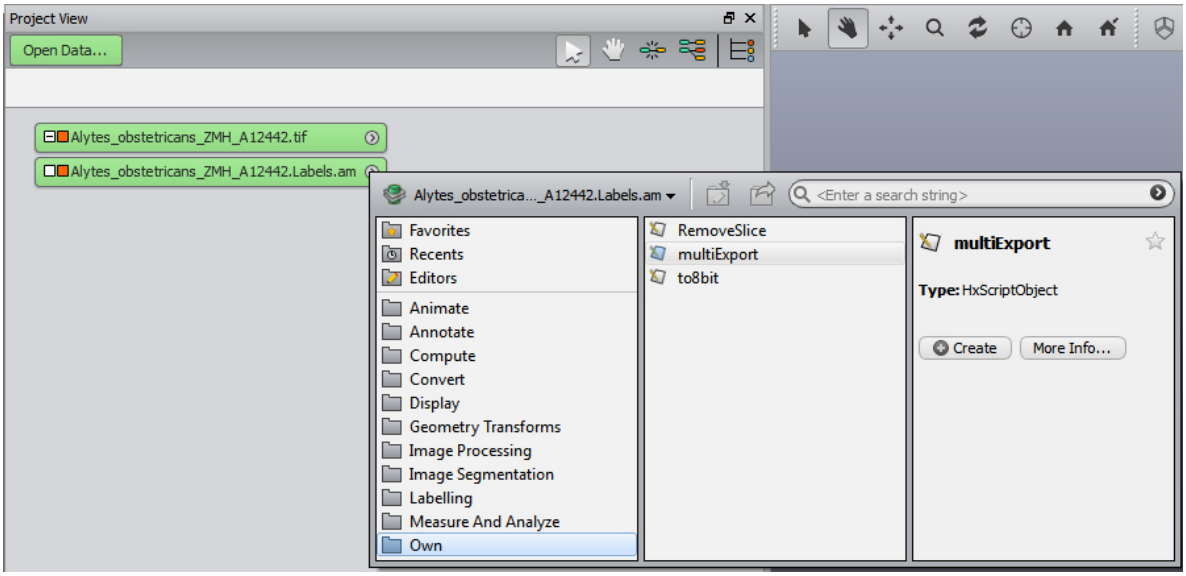


Figure 2: Path to *multiExport*.

How to call *multiExport*: Initially, the module has to be connected to a labels-object in first place (connection to an image stack is possible afterwards). Select the *LabelField*, open its menu and choose *MultiExport* in the folder *Own* (fig. 2).

What *multiExport* does: For each selected material *multiExport* performs the following steps (fig. 1):

1) *Arithmetic*: The *LabelField* is used as a mask to
    a. create a "black-and-white"-stack (*Result* in fig. 1), in which the voxel belonging to a certain *material* are given white color (1) and all others are set to black (0). This is automatically done when no *Imagedata* are connected. (Expression: $a == x$, where the *LabelField* is *Input A* and x is the *material* number)
    b. create a stack (*Result* in fig. 1), in which the voxel of a certain *material* are represented with their original gray values (as in the *Imagedata*), while all other voxels are set to black (0). This is automatically done when *Imagedata* are connected. (Expression: $(b == x)*a$, where *Imagedata* is *Input A*, the *LabelField* is *Input B*, and x is the *material* number)
2) If selected, the resulting stack (*Result* in fig. 1) is exported. If no surface should be exported, the iteration jumps back to the *Arithmetic* module and exports the next *material* until all selected *materials* are exported. If a surface should be exported, *multiExport* continues with step 3.
3) *Isosurface*: An *Isosurface* with the parameters *Threshold*, *Downsample x*, *y*, and *z* is created based on the *Result*. Note: Using a threshold is only meaningful if *Imagedata* are connected, otherwise all voxel belonging to the currently exported material are 1 and a threshold would select all or no voxel.
4) *Extract Surface*: A polygon surface (*ExtractedSurface* in fig. 1) is created based on the *Isosurface*. If *original* is selected for *obj* or *stl* the surface is exported as obj or stl file. When *simplified* is selected for *obj* or *stl multiExport* continues with step 5, if not it returns to 1.
5) *Cleanup*: The surface is cleaned using the commands *cleanup*, *removeDuplicatePoints*, *removeDuplicateTriangles*, and *removeDegenerateTriangles*.
6) *Smooth Surface*: The surface (*ExtractedSurface* in fig. 1) is smoothed using the parameters (*ItSmooth*, *Lambda*) provided at *Smooth*.
7) *Simplification Editor*: The polygon count is reduced to the percentage provided at *Reduction*.
8) *Iterations*: The surface is *ItTotal*-times cleaned, smoothed and reduced (as in steps 5-7). Then, the simplified surface is cleaned again and exported in the selected format.
9) Return to 1) to export next *material* or stop if all selected *materials* are exported.

Ports and parameters:

Labels	Port must be connected to a *LabelField*.
Imagedata	(Optional) Port can be connected to image data.
Directory	Port to choose the folder, in which the output files should be saved.
ImageStacks	Port to choose if image stacks containing the single materials should be exported. Export formats: *bmp, tiff, raw3d*. Note: the image stack (if provided) needs to be in 8bit (0-255) for bmp export. If it is not, it is converted which might be accompanied by image quality loss.
obj	Port to choose if surfaces in obj format should be exported. Choosing *original* will export the unchanged surfaces created by the *Extract Surface* module. Choosing *simplified* will export a smoothed and polygon count reduced surface.

stl	Port to choose if surfaces in stl format should be exported. Options as in *obj*.
Isosurface	Port to enter parameters for the *Isosurface* module. Check Amira® users guide or the help for more information.
ItTotal	Port to enter the total number of alternating smoothing and polygon count reduction steps.
Smooth	Port to enter the parameters for the *Smooth Surface* module. Check Amira® users guide or the help for more information.
Reduction	Port to enter the percentage (expressed between 0 and 1) of polygons that should be retained in each step of reduction. For example: if 0.9 is chosen, 90 % of the polygons are retained in each iteration; if a next iteration is performed (*ItTotal* > 1), 90 % of the 90 % are retained and so on.
Materials	Port to enter the number of the first and the last *material* that should be exported. Note: 0 is usually the *Exterior* and therefore excluded from this procedure.
Delete	Port to choose if all modules and objects created by *multiExport* should be deleted in the *Project View* after export.
Export	Port to start the export.

Final notes:
multiExport is provided "as is" and "with all faults." The authors cannot be held liable for any damages or data loss.
By using this script you agree to these conditions.

mutliExport has been developed and tested using Amira® versions 6.0.0 and 6.0.1 but it probably works with other versions as well. You are advised to save any data before using *multiExport*, as Amira® might crash during *multiExport* execution.

multiExport automatically creates a folder structure and assigns names to the exported files. If folders and files with identical names already exist in the selected output *Directory*, those are overwritten without warning.

[code of multiExport.rc]

```
# ######## IMPORTANT ############
# This script is provided "as is" and "with all faults". The authors cannot
# be held liable for any damages or data loss.
# By using this script you agree to these conditions.
#
# You are advised to save any data before using multiExport, as Amira(R)
# might crash during multiExport execution.
#
# This script is published as electronic supplementary material to the
# article
# K. Engelkes*, F. Friedrich, J. U. Hammel, A. Haas: A simple setup for
# episcopic microtomy and a digital image processing
# workflow to acquire high-quality volume data and 3D surface models of
small vertebrates. Zoomorphology.
# * Corresponding author: K. Engelkes, Center of Natural History (CeNak),
# Universität Hamburg, karolin.engelkes@uni-hamburg.de
# ################################

# Save this file and multiExport.scro in the folder "script-objects" of
# your Amira(R)-root-directory (e.g., C:/programs/share/script-objects/)
# and restart Amira(R).
```

```
# ######## Description ########
# This file includes multiExport as a module in Amira(R)
# Now you can access multiExport by right-clicking on a labels-object in
# the pool-window and selecting the own-folder. (Works for label-fields
# only, NOT for other image-data)

module -name "multiExport" \
      -primary "HxUniformLabelField3" \
      -category "Own" \
      -package "hxscriptobj" \
      -proc {
      set mod [[create HxScriptObject] setLabel multiExport]

        # include correct script (multiExport.scro) and execute
        "$mod" script setValue $AMIRA_ROOT/share/script-
        objects/multiExport.scro
        "$mod" fire

        # Connect Labels-Port with correct label data
            "$mod" Labels connect $PRIMARY
        "$mod" fire

       }
```

[code of multiExport.scro]

```
# Amira(R) Script
# Version of 02. June 2017
# ######## IMPORTANT ############
# This script is provided "as is" and "with all faults". The authors cannot
# be held liable for any damages or data loss.
# By using this script you agree to these conditions.
#
# You are advised to save any data before using multiExport, as Amira(R)
# might crash during multiExport execution.
#
# This script is published as electronic supplementary material to the
# article
# K. Engelkes*, F. Friedrich, J. U. Hammel, A. Haas: A simple setup for
# episcopic microtomy and a digital image processing
# workflow to acquire high-quality volume data and 3D surface models of
# small vertebrates. Zoomorphology.
# * Corresponding author: K. Engelkes, Center of Natural History (CeNak),
# Universität Hamburg, karolin.engelkes@uni-hamburg.de
# ################################
# Save this file and multiExport.rc in the folder "script-objects" of your
# Amira(R)-root-directory (e.g., C:/programs/share/script-objects/) and
# restart Amira(R).
# Now you can access multiExport by right-clicking on a labels-object in
# the pool-window (= Project View window) and selecting the own-folder.
# (Works for label-fields only, not for other image-data)
# ######## Description ########
# This script utilizes different functions that are already implemented in
# Amira(R) (e.g., Arithmetic, Isosurface, ...) to iteratively perform a
# selection of tasks listed in the following for each selected segmented
# material seperately
# (a) save an image stack in .tif-format.
# (b) save an image stack in .bmp-format. (the image-data have to be 8-bit;
# if they are not: use CastField before using multiExport or multiExport
# convertes the data, which might be incorporated with loss of image
# quality)
# (c) save an image stack in .raw-format.
# (d) save a surface in .obj- or .stl-format (surface is created using
# Isosurface, ExtractSurface).
# (e) save a reduced and smoothed version of the surface in .obj- or .stl-
# foramt.
# The resulting files are saved in subfolders of the selected directory.
# Attention: The script overwrites existing files without notification if
# they have the same name.
```

```
# ################################################
# ######### Proc constructor ####################
# ################################################
# This procedure is executed when the multiExport-module is created
"$this" proc constructor  { } {
      # Color for visualization in Pool-Window
      "$this" setIconColor {204 178 255}
      # Hide unnecessary Ports of ScriptObject
      "$this" script hide
      "$this" data hide
      # ### Create needed Ports ###############
      # Infoport for explanation
      "$this" newPortInfo "Info"
      "$this" "Info" setValue "Connect Labels-Field and optionally image
      data. \n (If no image data are connected the expression in the \n
      Arithmetic-Module will be a==number of material with a: Labels, \n
      else it will be (b==number of material)*a with a: Imagedata; b:
      Labels)"
      # Rename the built-in Port "data" (internally, the port is still
      # called "data", so data has to be used to refer to it)
      "$this" newPortConnection Labels HxUniformLabelField3
      # Port "Imagedata" [optional] for connection to original image data
      "$this" newPortConnection Imagedata HxUniformScalarField3
      # Infoport for explanation
      "$this" newPortInfo "Info1"
      "$this" "Info1" setValue "Select folder to store created image stacks
      and/or surfaces."
      # Port to choose directory for saveing
      "$this" newPortDirectory Directory
      # Infoport for explanation
      "$this" newPortInfo "Info2"
      "$this" "Info2" setValue "Choose desired output formats: (Important:
      bmp only supports \n 8 bit images. If images are not 8 bit they are
      converted to 8 bit. \n Note: during stl-exprot there might be a
      warning popup window \n that needs to be closed manually.) "
      # Port to choose Image stacks as output
      "$this" newPortToggleList "ImageStacks" 3
      "$this" ImageStacks setLabel 0 "BMP"
      "$this" ImageStacks setLabel 1 "2D Tiff"
      "$this" ImageStacks setLabel 2 "Raw Data 3D"
      # Port to choose obj-surfaces as output (original surface and/or
      # smoothed, reduced surface)
      "$this" newPortToggleList "obj" 2
      "$this" obj setLabel 0 "original"
      "$this" obj setLabel 1 "simplified"
      # Port to choose stl-surfaces as output (original surface and/or
      # smoothed, reduced surface)
      "$this" newPortToggleList "stl" 2
      "$this" stl setLabel 0 "original"
      "$this" stl setLabel 1 "simplified"
      # Infoport for explanation
      "$this" newPortInfo "Info3"
      "$this" "Info3" setValue "Optional: Parameter for Isosurface, \n
      reduction of polygon count amd smoothing of surface.\n Isosurface:
      Threshold and Downsample of Isosurface. \n ItTotal: Natural number
      for number of iterations of reduction and smoothing. \n Surface is
      cleaned in every iteration. \n Smooth: ItSmooth: natural number,
      Lambda: number between 0 and 1. \n See Smooth Surface module for
      meaning of ItSmooth and Lambda. \n Reduction: number between 0 and 1
      that gives percentage of polygon number \n that schould be retained.
      (e.g., 0.9 = 90% of all polygons retained in each iteration)"
      # Ports fo threshold and downsample of Isosurface
      "$this" newPortFloatTextN Isosurface 4
      "$this" Isosurface setLabel 0 "Threshold:"
      "$this" Isosurface setLabel 1 "Downsample x:"
      "$this" Isosurface setLabel 2 "y:"
      "$this" Isosurface setLabel 3 "z:"
      # Port for ItTotal
      "$this" newPortIntTextN ItTotal 1
      "$this" ItTotal setLabel 0 "ItTotal:"
      # Port for smoothing
```

```
      "$this" newPortFloatTextN Smooth 2
      # Number of iterations per smoothing step (natural number):
      "$this" Smooth setLabel 0 ItSmooth:
      # Factor lambda for smoothing (number between 0 and 1):
      "$this" Smooth setLabel 1 Lambda:
      # Port for reducion of polygon count: parameter between 0 and 1 that
      # is used to calculate the new number of polygons
      "$this" newPortFloatSlider Reduction
      "$this" Reduction setMinMax 0 1
      # Infoport for explanation
      "$this" newPortInfo "Info4"
      "$this" "Info4" setValue "Enter the number of the first and last
      material that should be exported:"
      # Port for first and last material to be exported
      "$this" newPortIntTextN Materials 2
      "$this" Materials setLabel 0 "First:"
      "$this" Materials setLabel 1 "Last:"
      # Port to choose if the modules, that are created during export,
      # should be deleted afterwards
      "$this" newPortToggleList "Delete" 1
      "$this" Delete setLabel 0 "delete during export created modules"
      # Button to start export
      "$this" newPortButtonList Export 1
      "$this" Export setLabel 0 "Go!"
      ######################################################################
      # ############## Set default parameters for some ports ##############
      # ### Values for Isosurface
      "$this" Isosurface setValue 0 0
      "$this" Isosurface setValue 1 1
      "$this" Isosurface setValue 2 1
      "$this" Isosurface setValue 3 1
      # ### Parameter for ItTotal
      "$this" ItTotal setValue 0 5
      # ### Parameter for Smoothing
      "$this" Smooth setMinMax 1 0 1
      "$this" Smooth setValues 4 0.6
      # ### Parameter for reduction of polygon count
      "$this" Reduction setValue 0.9
      # ### Default output directory
      "$this" Directory setValue "C:/Amira_output"
      # ### Set default value for port Delete
      "$this" Delete setValue 0 1
}
# ###################################
# ####### Procedure compute ########
# ###################################
# This procedure is always executed when any changes are made within the
# properties window (e.g., when the "Go!"-Button is hit).
"$this" proc compute { } {
      # ### Set some default values and directory for saving
      # Get name of LabelField
      set myLabels ["$this" Labels source]
      if {"$myLabels" != ""} {
            # Get value of Port "Directory" (= folder, in which the output
      # data will be saved)
            set dir ["$this" Directory getValue]
            if {("$dir" == "") || ("$dir"=="C:/Amira_output") } {
                  # If no output-directory is set: read directory, in which
                  # the LabelField is stored and use it as default
                  # directory
                  set dir "[$myLabels parameters Filename getValue]_out"
                  if {"$dir" != "Filename: unknown command_out"} {
                        "$this" Directory setValue $dir
                  }
            }
            # ### Default values for first and last material that should be
            # exported
            # Make list with 2 entries: 1. entry is 0 (for material
            # exterior), 2. entry is number of last material (= total
            # number of materials)
```

```
            set range [$myLabels getRange]
            set max [lindex $range 1]
            "$this" Materials setMinMaxAll 0 $max 0 $max
            if {["$this" Materials getValue 1] == "0"} {
                  "$this" Materials setValue 0 1
                  "$this" Materials setValue 1 $max
            }

            # Read first and last material that will be exported from
            # properties window and store in variables
            # First material
            set i ["$this" Materials getValue 0]
            # Last material
            set NumberOfMaterials ["$this" Materials getValue 1]
            # ############################################################
            # ###### Export materials if and only if "Go!" was hit ########
            if {["$this" Export wasHit]==1} {
                  ####### Set variables (read input values from porperties
                  # window and store in variables) #######
                  # Should the original surface be exported as obj? no=0;
                  # yes=1
                  set saveobj ["$this" obj getValue 0]
                  # Should the reduced and smoothed surface be exported as
                  # obj? no=0; yes=1
                  set saveobjp ["$this" obj getValue 1]
                  # Should the original surface be exported as stl? no=0;
                  # yes=1
                  set savestl ["$this" stl getValue 0]
                  # Should the reduced and smoothed surface be exported as
                  # stl? no=0; yes=1
                  set savestlp ["$this" stl getValue 1]
                  # Should a .bmp-image-stack be exported? no=0; yes=1
                  set savebmp ["$this" ImageStacks getValue 0]
                  # Should a .tif-image-stack be exported? no=0; yes=1
                  set savetif ["$this" ImageStacks getValue 1]
                  # Should a RAW-3D-image-stack be exported? no=0; yes=1
                  set saveraw ["$this" ImageStacks getValue 2]
                  # Total number of iterations of reducing and smoothing
                  set ItTotal ["$this" ItTotal getValue 0]
                  # Parameter for reduction of polygon count
                  set reduceTo ["$this" Reduction getValue]
                  # Number of iterations for smoothing
                  set ItSmooth ["$this" Smooth getValue 0]
                  # Factor Lambda for smoothing
                  set lambda ["$this" Smooth getValue 1]
###########################################################################
                  # ######## Create Arithmetic-module and set values ######
                  set arithmetic [[create HxArithmetic] setLabel
                  myArithmetic]
                  # If multiExport is connected to image data: set image
                  # data as InputA and Labels as InputB; else set Labels as
                  # InputA
                  set Imagedata ["$this" Imagedata getValue]
                  if {"$Imagedata" != ""} {
                        $arithmetic inputA connect $Imagedata
                        $arithmetic inputB connect $myLabels
                  } else {
                        $arithmetic inputA connect $myLabels
                  }
###########################################################################
                  # ### If surfaces should be exported: execute Arithmetic
                  # and create an Isosurface module that is connected to
                  # the result of Arithmetic
                  if {["$this" obj getValue 0] == 1 || ["$this" obj
                  getValue 1] == 1 || ["$this" stl getValue 0] == 1 ||
                  ["$this" stl getValue 1] == 1} {
                        # No idea why, but this is necessary to execute
                        # "Arithmetic fire"
                        $arithmetic doIt setValue 0
                        $arithmetic doIt setValue 0
                        # Execute Arithmetic
```

```
                    $arithmetic fire
                    # Store result of Arithmetic in variable
                    set myRes [$arithmetic getResult]
                    # Threshold and downsample parameters for
                    # sosurface
                    set Threshold ["$this" Isosurface getValue 0]
                    set DownsampleX ["$this" Isosurface getValue 1]
                    set DownsampleY ["$this" Isosurface getValue 2]
                    set DownsampleZ ["$this" Isosurface getValue 3]
                    # Create Isosurface-module, connect to result of
                    # Aritmethic and set default threshold (user
                    # defined threshold is set below) and user-selected
                    # downsample parameters
                    set myIsosurf [[create HxIsosurface] setLabel
                    myIsosurface]
                    "$myIsosurf" data connect $myRes
                    set max ["$myIsosurf" threshold getMaxValue]
                    "$myIsosurf" threshold setValue 2
                    set max ["$myIsosurf" threshold getMaxValue]
                    if {$DownsampleX == 1 && $DownsampleY == 1 &&
                    $DownsampleY == 1} {
                          "$myIsosurf" options setState 1 0
                    } else {
                          "$myIsosurf" options setState 1 1
                          "$myIsosurf" resolution setValue 0
                          $DownsampleX
                          "$myIsosurf" resolution setValue 1
                          $DownsampleY
                          "$myIsosurf" resolution setValue 2
                          $DownsampleZ
                    }
              }
#########################################################################
              ############### Initialize time measuring, variables,
              # modules, outout folder, ... ###############
              # Variable for feedback of currently processed material
              # in console
              set j 1
              # Total number of materials that should be exported
              set numberges [expr $NumberOfMaterials-$i+1]
              # Display feedback in console
              echo "Materials to be exported: $i to $NumberOfMaterials
              ($numberges materials)."
              # Variable total time
              set tges 0
              # ### Expand module-network as needed
              # If needed: create SurfaceExtractor and connect to
              # Isosurface
              if {$saveobj==1 || $saveobjp==1 || $savestl==1 ||
                    $savestlp==1} {
                    set SurfaceExtractor [[create HxViewBaseExtract]
                    setLabel myExtractSurface]
                    "$SurfaceExtractor" module1 connect $myIsosurf
              }
              # ### Test, if output folders exists; if not: create them
              # xxx_out-folder
              if {[file exists "$dir"] == 0 } {
                    file mkdir $dir
                    echo "File '$dir' created."
              }
              # ### Optional output-folder for image stacks or
              # surfaces:
              # obj
              if {[file exists "$dir/surfaces_obj"] == 0 &&
                    $saveobj==1} {
                    file mkdir $dir/surfaces_obj
                    echo "File '$dir/surfaces_obj' created."
              }
              if {[file exists "$dir/simplified_surfaces_obj"] == 0 &&
                    $saveobjp==1} {
                    file mkdir $dir/simplified_surfaces_obj
```

```
                echo "File '$dir/simplified_surfaces_obj' created."
            }
            # stl
            if {[file exists "$dir/surfaces_stl"] == 0 &&
                $savestl==1} {
                file mkdir $dir/surfaces_stl
                echo "File '$dir/surfaces_stl' created."
            }
            if {[file exists "$dir/simplified_surfaces_stl"] == 0 &&
                $savestlp==1} {
                file mkdir $dir/simplified_surfaces_stl
                echo "File '$dir/simplified_surfaces_stl' created."
            }
            # image stacks
            if {[file exists "$dir/BMPs"] == 0 && $savebmp==1} {
                file mkdir $dir/BMPs
                echo "File '$dir/BMPs' created."
            }
            if {[file exists "$dir/TIFs"] == 0 && $savetif==1} {
                file mkdir $dir/TIFs
                echo "File '$dir/TIFs' created."
            }
            if {[file exists "$dir/RAW3Ds"] == 0 && $saveraw==1} {
                file mkdir $dir/RAW3Ds
                echo "File '$dir/RAW3Ds' created."
            }
##########################################################
            # ############## Execute multiExport
            # Iteration for all material numbers selected
            while {$i <= $NumberOfMaterials} {
                # Time the present iteration started
                set timeIn [clock seconds]
                # Variable to store present material name
                set material [$myLabels getMaterialName $i]
                # ### Arithmetic ######
                echo "---------- New material ($j/$numberges;
                Number of Material: $i; Time passed: $tges s) -----
                -----------------"
                # Set new expression in Arithmetic (different
                # expressions for image data connected and not
                # connected)
                if {"$Imagedata" != ""} {
                    $arithmetic expr setValue "(b==$i)*a"
                } else {
                    $arithmetic expr setValue "a==$i"
                }
                $arithmetic doIt setValue 0
                # Execute Arithmetic (same as hit "Apply")
                $arithmetic fire
                # Store result name in variable
                set myRes [$arithmetic getResult]
                # ### If selected: save as .bmp ####
                if {$savebmp==1} {
                    file mkdir "$dir/BMPs/$i-$material"
                    # If result not 8-bit: convert to 8-bit and
                    save; else save directly
                    $myRes select
                    set test [$myRes Datainfo getValue]
                    $myRes deselect
                    if {[lindex $test 1] != "8-bit"} {
                        # Create CastField and set parameters
                        set CastField [[create HxCastField]
                        setLabel myCastField]
                        $CastField data connect $myRes
                        $CastField outputType setValue 0
                        $CastField scaling  setValue 0 0.00
                        # Execute CastField
                        $CastField action setValue 0
                        $CastField fire
                        # Save
                        set CastRes [$CastField getResult]
```

```
                $CastRes exportData BMP "$dir/BMPs/$i-
                $material/$i-$material.####.bmp"
                # Cleanup
                remove $CastRes
                remove $CastField
        } else {
                $myRes exportData BMP "$dir/BMPs/$i-
                $material/$i-$material.####.bmp"
        }
}
# ### If selected: save as .tif  ######
if {$savetif==1} {
        file mkdir $dir/TIFs/$i-$material
        $myRes exportData "2D TIF" "$dir/TIFs/$i-
        $material/$i-$material-####.tif"
}
# ### If selected: save as .raw ####
if {$saveraw==1} {
        $myRes exportData "RAW Data 3D"
        "$dir/RAW3Ds/$i-$material.raw"
}
# ### If selected: create surface ####
if {$saveobj==1 || $saveobjp==1 || $savestl==1 ||
        $savestlp==1} {
        # Set user input for threshold and execute
        # Isosurface
        "$myIsosurf" threshold setValue $Threshold
        "$myIsosurf" doIt setValue 0
        "$myIsosurf" fire
        # Execute ExtractSurface und store result
        # name
        $SurfaceExtractor action setValue 0
        $SurfaceExtractor fire
        set surface [$SurfaceExtractor getResult]
        # If selected: save original surface as .obj
        if {$saveobj==1} {
                $surface exportData Wavefront
                $dir/surfaces_obj/$i-$material.obj
                echo "$dir/$i-$material $i.obj saved."
        }
        # If selected: save original surface as .stl
        if {$savestl==1} {
                $surface exportData STL
                $dir/surfaces_stl/$i-$material.stl
                echo "$dir/$i-$material $i.stl saved."
        }
        # ### If selected: reduce and smooth surface
        if {($saveobjp==1) || ($savestlp==1)} {
                set k 0
                while {$k < $ItTotal} {
                        # Clean surface
                        $surface cleanup
                        $surface removeDuplicatePoints
                        $surface removeDuplicateTriangles
                        $surface removeDegenerateTriangles
                        # Create SmoothSurface module, set
                        # values, and smooth surface
                        set SmoothSurface [[create
                        HxSurfaceSmooth] setLabel
                        mySmoothSurface]
                        $SmoothSurface data connect
                        $surface
                        $SmoothSurface parameters setValue
                        $ItSmooth
                        $SmoothSurface parameters setValue
                        1 $lambda
                        $SmoothSurface action setValue 0
                        $SmoothSurface fire
                        set surfaceRes [$SmoothSurface
                        getResult]
                        remove "$surface"
```

```
                                    set surface $surfaceRes
                                    # Get present number of triangles
                                    # and store as n
                                    set n [$surface getNumTriangles]
                                    # Reduce polygon count to
                                    $reduceTo*100 %:
                                    set Simplifier [[create
                                    HxSimplifier] setLabel
                                    mySimplifier]
                                    $Simplifier attach $surface
                                    $Simplifier simplifyParameters
                                    setValue [expr $n*$reduceTo]
                                    $Simplifier simplifyAction
                                    setValue 0
                                    $Simplifier fire
                                    set k [expr $k + 1]
                              }
                              # Final smoothing
                              set SmoothSurface [[create
                                    HxSurfaceSmooth] setLabel
                                    mySmoothSurface]
                                    $SmoothSurface data connect
                                    "$surface"
                                    $SmoothSurface parameters setValue
                                    0 $ItSmooth
                                    $SmoothSurface parameters setValue
                                    1 $lambda
                                    $SmoothSurface action setValue 0
                                    $SmoothSurface fire
                              # Clean surface
                              "$surface" cleanup
                              "$surface" removeDuplicatePoints
                              "$surface" removeDuplicateTriangles
                              "$surface" removeDegenerateTriangles
                              set surfaceRes [$SmoothSurface
                              getResult]
                              remove "$surface"
                              set surface $surfaceRes
                              # ###Save simplified surface
                              # obj
                              if {($saveobjp==1)} {
                                    "$surface" exportData Wavefront
                                    $dir/simplified_surfaces_obj/$i-
                                    $material.processed.obj
                                    set surface ["$surface" setLabel
                                    MySmoothed]
                              }
                              # stl
                              if {($savestlp==1)} {
                                    "$surface" exportData STL
                                    $dir/simplified_surfaces_stl/$i-
                                    $material.processed.stl
                                    set surface ["$surface" setLabel
                                    MySmoothed]
                              }
                        }
                        remove "$surface"
                  }
                  # Calculate passed time and diplay feedback in
                  # console
                  set timeOut [clock seconds]
                  set timeCalc [expr $timeOut-$timeIn]
                  set tges [expr $tges+$timeCalc]
                  echo "Material $i done. Calculation time: $timeCalc
                  s."
                  # Counters plus 1 for next material
                  set i [expr $i+1]
                  set j [expr $j+1]
                  # Test, if export should be interrupted
                  "$this" testBreak
            }
```

```
                    # Delete the list of materials that should be exported
                    # (if not deleted, there might be problems with exporting
                    # all materials if multiExport is run again)
                    if {[info exists limits]==1} {
                          unset limits
                    }
                    # ### Feedback in console that export is finished #####
                    echo "DONE!!! Total time: $tges s (= [expr $tges/60.0]
                    min)"
                    # Delete created modules if selected
                    set delete ["$this" Delete getValue 0]
                    if {$delete == 1} {
                          remove "$arithmetic"
                          remove "$myRes"
                          if {[info exists surface] == 1} {
                                remove "$surface"
                          }
                          if {[info exists myIsosurf] == 1} {
                                remove "$myIsosurf"
                          }
                          if {[info exists SurfaceExtractor] == 1} {
                                remove "$SurfaceExtractor"
                          }
                    }
              }
        }
}
```

Online Resource 3: Dscriptive statistics of the sets of shortest distances between the observed midpoints and the respectively fitted lines.
Min: minimum distance in the distance set; max: maximum distance; SD: standard deviation. Unit: pixel.

alignment / spot group	1	2	3	4	5
i	min = 2.7170 max = 84.4632 median = 27.7751 mean = 32.3219 SD = 18.0149	min = 1.9951 max = 77.0898 median = 25.2942 mean = 29.4086 SD = 16.2500	min = 0.3110 max = 16.7097 median = 4.2517 mean = 4.8341 SD = 3.5168	min = 0.7080 max = 31.0147 median = 10.7734 mean = 13.2303 SD = 7.8628	min = 0.1920 max = 11.4705 median = 6.2054 mean = 5.7771 SD = 2.5522
ii	min = 10.7082 max = 119.4025 median = 43.3075 mean = 43.8794 SD = 20.2944	min = 8.0510 max = 108.2030 median = 37.8241 mean = 39.8440 SD = 18.7713	min = 0.3802 max = 28.2773 median = 9.2485 mean = 10.0947 SD = 5.5466	min = 1.4641 max = 33.9496 median = 13.4421 mean = 14.7350 SD = 7.4305	min = 0.7397 max = 16.0170 median = 7.7151 mean = 7.8824 SD = 3.2105
iii	min = 0.3322 max = 44.3379 median = 8.4745 mean = 11.0237 SD = 7.2977	min = 0.1497 max = 41.1318 median = 8.0617 mean = 10.0999 SD = 6.5527	min = 1.3058 max = 7.5706 median = 4.5397 mean = 4.40561 SD = 1.3077	min = 0.7702 max = 21.4971 median = 5.3330 mean = 6.5919 SD = 4.5394	min = 0.3790 max = 5.4968 median = 2.7452 mean = 2.9265 SD = 1.2252
iv	min = 1.2534 max = 4.4496 median = 6.4047 mean = 7.1108 SD = 3.0712	min = 0.2 max = 15.9036 median = 6.3519 mean = 6.7962 SD = 3.1171	min = 0.1970 max = 9.2109 median = 3.2612 mean = 3.7257 SD = 2.1699	min = 0.1475 max = 8.4655 median = 2.9071 mean = 3.1748 SD = 1.8012	min = 0.1692 max = 8.1152 median = 2.4398 mean = 2.7323 SD = 1.6313

Online Resource 4: Results of Kruskal-Wallis rank sum test and Dunn's test on distance sets

Kruskal-Wallis test (function *kruskal.test* as implemented in R-package *stats*)
$\chi^2 = 3375.8$, df = 19, p-value < 2.2e-16
Dunn's test with p-value adjustment according to Holm (function *posthoc.kruskal.dunn.test* in R-package *PMCMR*)
First value: test statistic; second value: p-value. All values rounded to four decimals. Green: No significant difference; yellow: significant difference. Values smaller than 0.05 were considered significant.

	1.i	2.i	3.i	4.i	5.i	1.ii	2.ii	3.ii	4.ii	5.ii	1.iii	2.iii	3.iii	4.iii	5.iii	1.iv	2.iv	3.iv	4.iv
2.i	0.7506 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3.i	21.4675 0	20.7169 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4.i	9.5582 0	8.8076 0	11.9093 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5.i	18.3878 0	17.6372 0	3.0797 0.0767	8.8296 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1.ii	3.7027 0.0096	4.5154 0.0003	26.9452 0	14.0512 0	23.6109 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2.ii	2.8355 0.1418	3.6482 0.0114	26.0780 0	13.1840 0	22.7436 0	0.9532 1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3.ii	13.1648 0	12.3521 0	10.0771 0	2.8163 0.1457	6.7434 0	18.5390 0	17.5858 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4.ii	8.1468 0	7.3342 0	15.0956 0	2.2016 0.6111	11.7613 0	13.0238 0	12.0706 0	5.5152 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5.ii	15.4567 0	14.6441 0	7.7857 0	5.1083 0	4.4514 0.0004	21.0581 0	20.1049 0	2.5191 0.3294	8.0343 0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
1.iii	10.6048 0	9.9462 0	8.2323 0	2.2178 0.6111	5.5300 0	14.4472 0	13.7009 0	0.0687 1	4.2497 0.0011	2.0412 0.8659	NA	NA	NA	NA	NA	NA	NA	NA	NA
2.iii	11.3594 0	10.7008 0	7.4777 0	2.9724 0.1064	4.7753 0.0001	15.2486 0	14.5022 0	0.7326 1	5.0510 0	1.2399 1	0.6804 1	NA	NA	NA	NA	NA	NA	NA	NA
3.iii	19.0589 0	18.4002 0	0.2218 1	10.6719 0	2.9241 0.1140	23.4242 0	22.6779 0	8.9083 0	13.2266 0	6.9358 0	7.6226 0	6.9422 0	NA	NA	NA	NA	NA	NA	NA
4.iii	15.8939 0	15.2353 0	2.9432 0.1105	7.5069 0	0.2408 1	20.0635 0	19.3172 0	5.5476 0	9.8659 0	3.5751 0.0144	4.7690 0.0001	4.0885 0.0021	2.8537 0.1383	NA	NA	NA	NA	NA	NA
5.iii	22.6985 0	22.0399 0	3.8614 0.0053	14.3115 0	6.5638 0	27.2890 0	26.5427 0	12.7731 0	17.0914 0	10.8006 0	10.9044 0	10.2239 0	3.2817 0.0402	6.1354 0	NA	NA	NA	NA	NA
1.iv	14.5756 0	13.8853 0	5.1677 0	5.7851 0	2.3353 0.5077	18.9363 0	18.1491 0	3.6255 0.0121	8.1804 0	1.5451 1	3.0920 0.0756	2.3853 0.4608	4.8248 0.0001	1.8610 1	8.2331 0	NA	NA	NA	NA
2.iv	15.2226 0	14.5322 0	4.5207 0.0003	6.4321 0	1.6884 1	19.6276 0	18.8404 0	4.3168 0.0008	8.8717 0	2.23639 0.6078	3.6700 0.0107	2.9633 0.1065	4.2468 0.0011	1.2830 1	7.6551 0	0.6022 1	NA	NA	NA
3.iv	21.7705 0	21.0801 0	2.0272 0.8659	12.9800 0	4.8595 0.0001	26.6246 0	25.8374 0	11.3138 0	15.8687 0	9.2334 0	9.5202 0	8.8136 0	1.6035 1	4.5673 0.0003	1.8048 1	6.69703 0	6.0949 0	NA	NA
4.iv	23.1084 0	22.4181 0	3.3651 0.0306	14.3179 0	6.1975 0	28.0544 0	27.2672 0	12.7436 0	17.2984 0	10.6631 0	10.7157 0	10.0090 0	2.7989 0.1487	5.7627 0	0.6094 1	7.9424 0	7.3403 0	1.2454 1	NA
5.iv	24.2096 0	23.5193 0	4.4663 0.0004	15.4191 0	7.2986 0	29.2311 0	28.4439 0	13.9203 0	18.4751 0	11.8398 0	11.6995 0	10.9928 0	3.7828 0.0071	6.7465 0	0.3744 1	8.9674 0	8.3652 0	2.2704 0.5796	1.0250 1

Supporting Information Chapter six

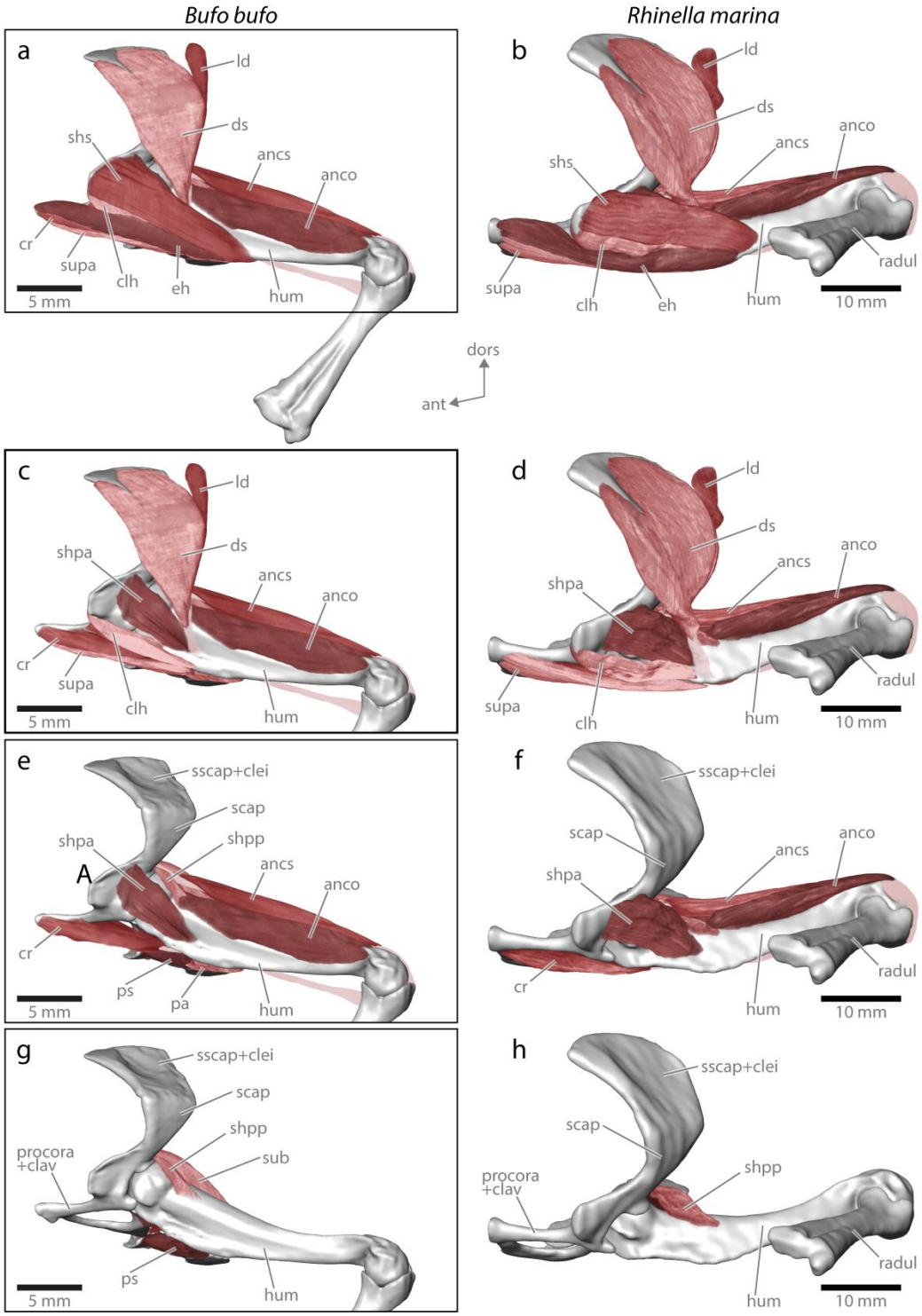


Fig. S1 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**, **g**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**, **h**). Left side only, anterolateral views, muscle layers successively removed. Combination of surfaces (skeleton) and volume renders (muscles), derived from µCT-volumes. Red: muscle (different shades for better visual separation of adjacent muscles); light gray: skeletal element with no distinction of bone and cartilage. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; ant: anterior; clav: clavicula; clei: cleithrum; clh: m. cleidohumeralis; cr: m. coracoradialis; dors: dorsal; ds: m. dorsalis scapulae; eh: m. episternohumeralis; hum: humerus; ld: m. latissimus dorsi; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; radul: radioulna; scap: scapula; shs: m. scapulohumeralis

superficialis; shpa: m. scapulohumeralis profundus anterior; shpp: m. scapulohumeralis profundus posterior; sscap: suprascapula; supa: m. supracoracoideus portio anterior; sub: m. subcoracoscapularis.

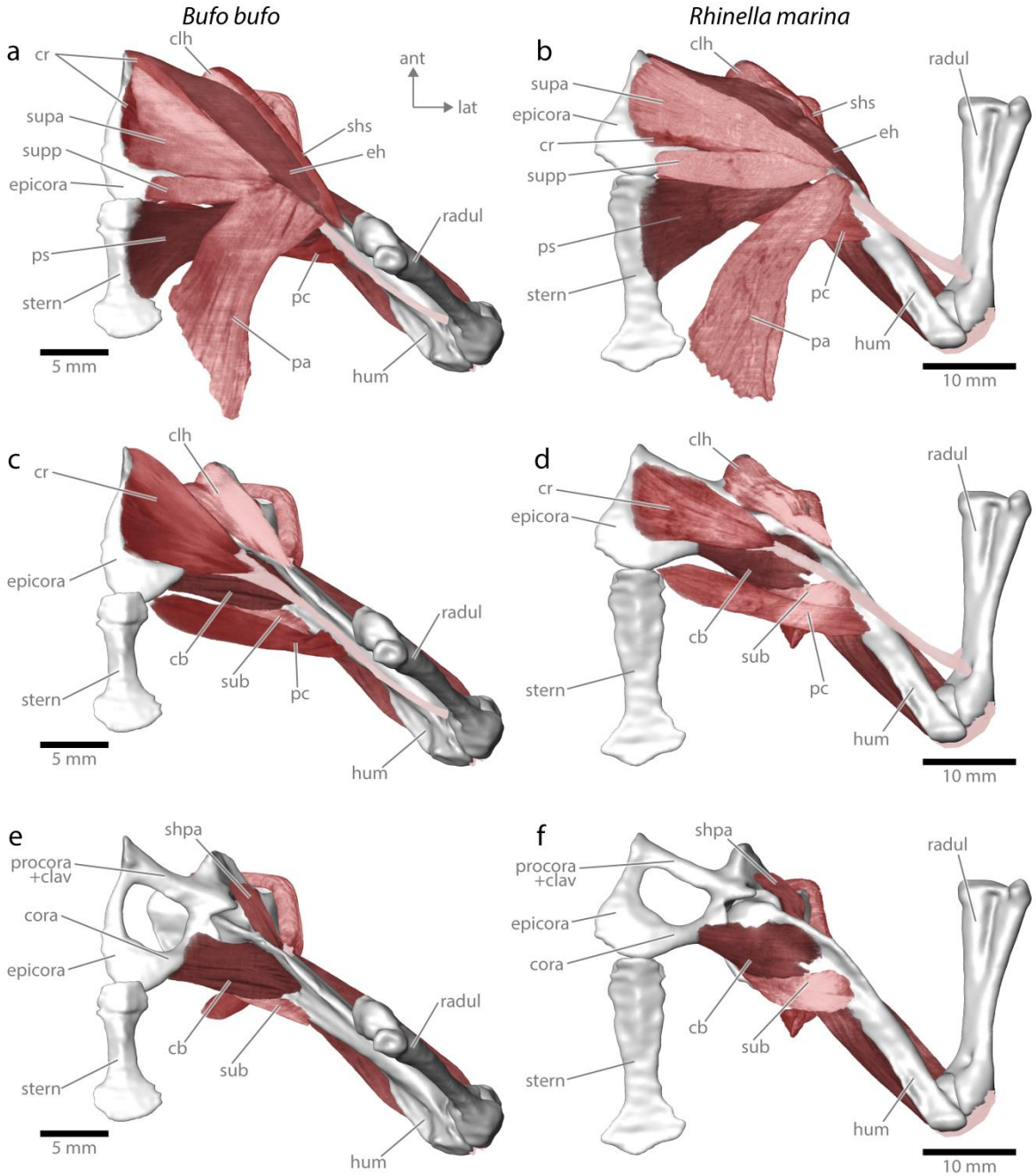


Fig. S2 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**). Left side only, ventral views, muscle layers successively removed. Combination of surfaces (skeleton) and volume renders (muscles), derived from µCT-volumes. Red: muscle (different shades for better visual separation of adjacent muscles); light gray: skeletal element with no distinction of bone and cartilage. ant: anterior; cb: m. coracobrachialis; clav: clavicula; clh: m. cleidohumeralis; cora: coracoid; cr: m. coracoradialis; eh: m. episternohumeralis; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; procora: procoracoid cartilage; ps: m. pectoralis portio sternalis; radul: radioulna; shs: m. scapulohumeralis superficialis; shpa: m. scapulohumeralis profundus anterior; supa: m. supracoracoideus portio anterior; supp: m. supracoracoideus portio posterior; stern: sternum; sub: m. subcoracoscapularis.

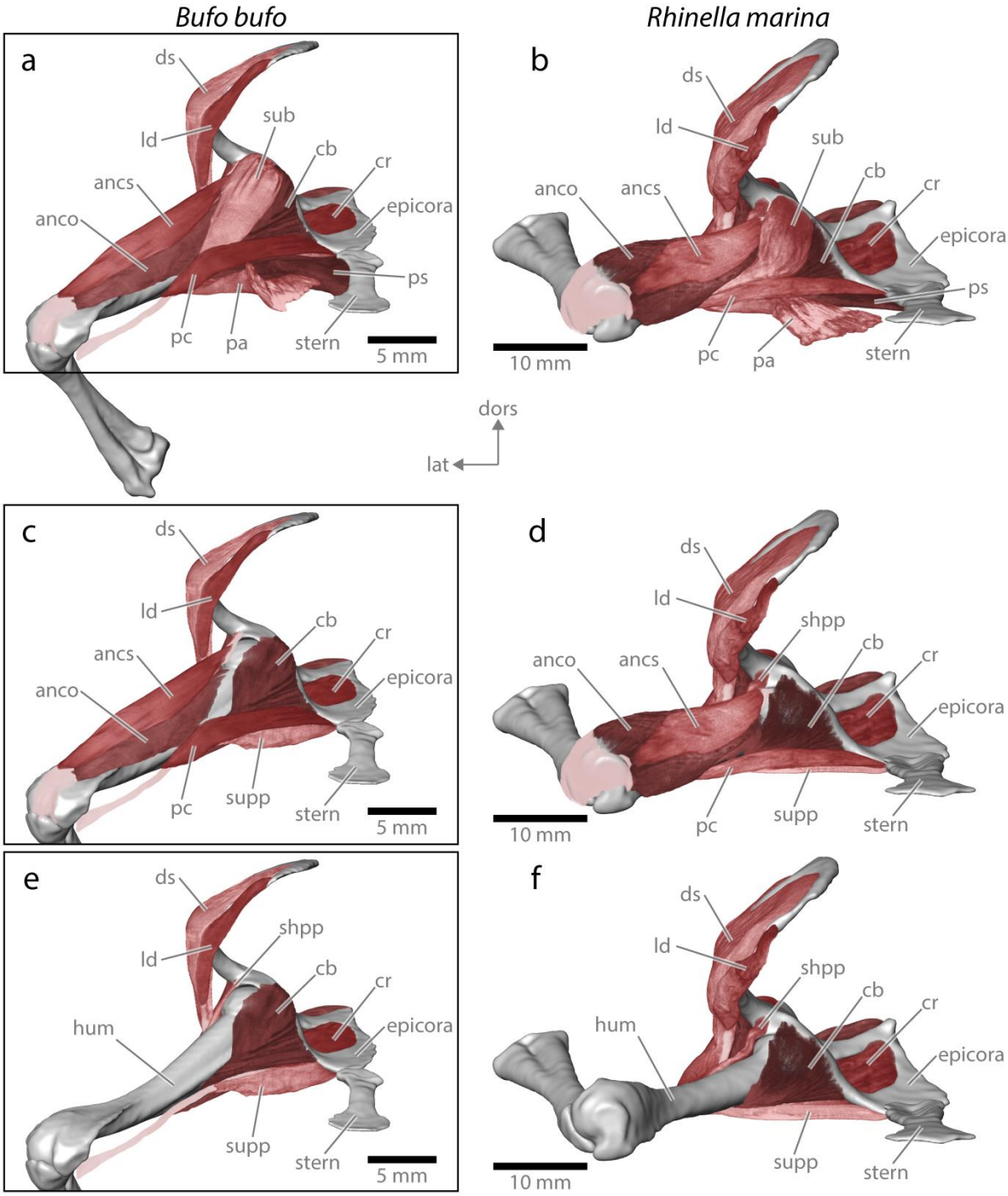


Fig. S3 Shoulder joint muscles in *Bufo bufo* (ZMH A04664, left, **a**, **c**, **e**) and *Rhinella marina* (ZMH A15443, right, **b**, **d**, **f**). Left side only, posterior views, muscle layers successively removed. Combination of surfaces (skeleton) and volume renders (muscles), derived from µCT-volumes. Red: muscle (different shades for better visual separation of adjacent muscles); light gray: skeletal element with no distinction of bone and cartilage. anco: heads of m. anconaeus not crossing the shoulder joint; ancs: m. anconaeus caput scapulare; ant: anterior; cb: m. coracobrachialis; cr: m. coracoradialis; ds: m. dorsalis scapulae; epicora: epicoracoid cartilage; hum: humerus; lat: lateral; ld: m. latissimus dorsi; pa: m. pectoralis portio abdominalis; pc: m. pectoralis portio coracoidea; ps: m. pectoralis portio sternalis; shpp: m. scapulohumeralis profundus posterior; stern: sternum; sub: m. subcoracoscapularis; supp: m. supracoracoideus portio posterior.