Genomic architecture in the *Daphnia longispina* species complex

Dissertation with the aim of achieving a doctoral degree

at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Biology

Universität Hamburg

Submitted by

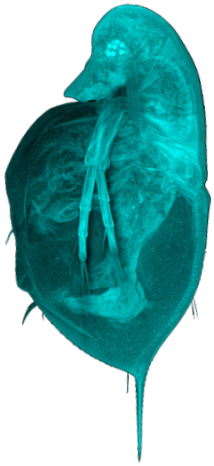
Jana Helene Nickel

Hamburg, September 2023

First Examiner: Dr. Mathilde Cordellier

Second Examiner: Prof. Dr. Susanne Dobler

Date of oral defense: 15.12.2023



Contents

Summary

Hybridization between species across the tree of life is an important evolutionary force that shapes the species' genomes. The short- and long-term evolutionary consequences of hybridization vary among taxa and are affected by neutral processes, recombination, or selection. The impact of these processes can vary between populations and individuals and lead to diverse outcomes, such as the erosion of reproductive barriers or the introduction of new genetic material for adaptation. The overarching questions addressed in this thesis are about the causes and frequency of recent hybridization in hybrid zones, and how genomic patterns vary among populations and species. I examine these questions in taxa of the zooplankton *Daphnia longispina* species complex, which is known to hybridize frequently and form freshwater mosaic hybrid zones. To facilitate this, I apply, adapt, and develop appropriate genomic methods to investigate genome-wide hybridization and introgression.

Daphnia resting eggs accumulate over time in freshwater ecosystems and are particularly well suited for population genomic studies of hybrid zones and the temporal changes they experience. However, methods were hindered by small amounts of potentially degraded DNA. In **Chapter 2**, I introduce and test a cost-saving method for whole-genome sequencing of *Daphnia* resting eggs from the resting egg bank using whole genome amplification and a pre-sequencing contamination screening. This allows us, for the first time, to directly investigate genomic diversity and hybridization in *Daphnia* populations over several decades.

Previous studies of the *Daphnia longispina* species complex have found frequent hybridization, but lacked a reference genome to characterize genome-wide variability. The *D. galeata* reference genome I present in **Chapter 3**, allowed us to identify extensive hybridization and introgression in ecologically diverse habitats. In total, we resequenced and analyzed 49 genomes from the three species and their hybrids, from genotypes sampled in the water column and from single resting eggs extracted from sediment cores. Introgression patterns revealed a complex history reflecting multiple generations of hybridization and backcrossing.

Building on this work, I studied the hybridization dynamics of two populations in more detail in **Chapter 4**. I further improved the reference genome to a chromosome scale and generated more extensive time-series genomic data. Using a novel junctions approach based on how recombination breaks up genomic blocks of contiguous ancestry over time, I demonstrate how it can be used to estimate the extent and timing of hybridization events. I identified recurrent hybridization across the genome over several decades in both populations and diverse genome-wide hybrid ancestry among individuals. These findings establish the *Daphnia longispina* species complex as a hybrid system with unique features, such as continuous

hybridization between multiple species in a mosaic hybrid zone, not previously described in a natural hybrid population.

These temporal hybridization dynamics are important to consider as changes to hybrid zones are accelerated by anthropogenic changes in *Daphnia* and other hybrid systems. Ultimately, I show the temporal dynamics and the diverse genomic landscapes of hybrid individuals with unprecedented resolution. This thesis illustrates the evolutionary complexity required to understand all factors that shape mosaic hybrid zones, such as spatial and temporal changes, recurrent hybridization, and the presence of multiple species.

Zusammenfassung

Die Hybridisierung zwischen Arten ist bei allen Lebewesen ein wichtiger evolutionärer Faktor, der das Genom von Arten formt. Die kurz- und langfristigen evolutionären Folgen der Hybridisierung variieren je nach Taxa und werden durch neutrale Prozesse, Rekombination oder Selektion beeinflusst. Die Auswirkungen dieser Prozesse können sich von Population zu Population und von Individuum zu Individuum unterscheiden und zu vielfältigen Auswirkungen führen, wie z.B. das Auflösen von reproduktiven Isolationsmechanismen oder der Austausch von neuer genetischer Variation, die zur Adaptation beitragen kann. Die übergreifenden Fragen, die in dieser Doktorarbeit behandelt werden, sind die Ursachen und Verbreitung von zeitgenössischer Hybridisierung in Hybridzonen und wie sich die Muster im Genom zwischen Populationen und Arten unterscheiden. Ich untersuche diese Fragen am Beispiel des *Daphnia longispina* Artenkomplexes, eine Gruppe von Süßwasser-Zooplankton, von dem bekannt ist, dass sie häufig hybridisieren und Mosaik-Hybridzonen bilden. Um dies zu ermöglichen, wende ich geeignete genomische Methoden an, passe sie an und entwickele sie weiter, um genomweiten Hybridisierung und Introgression zu untersuchen.

Daphnia Dauereier sammeln sich im Laufe der Zeit in Süßwasser-Ökosystemen an und eignen sich damit besonders gut für populationsgenomische Untersuchungen von Hybridzonen und deren zeitliche Veränderungen. Die bisherigen Methoden wurden allerdings durch geringe Mengen potenziell degradierter DNA erschwert. In **Kapitel 2** stelle ich eine kostengünstige Methode vor, um ganze Genome von *Daphnia* Dauereiern aus der Dauereierbank zu sequenzieren vor und teste sie, wobei Genomamplifikation und ein Kontaminations-Screening verwendet werden. Dies erlaubt uns erstmalig, die genomische Diversität und Hybridisierung in *Daphnia*-Populationen über mehrere Jahrzehnte hinweg direkt zu untersuchen.

Frühere Studien des *Daphnia longispina* Artenkomplexes haben häufige Hybridisierung gefunden, aber es fehlte ein Referenzgenom, um die genomweite Variabilität zu charakterisieren. Mit dem *D. galeata* Referenzgenom, das ich in **Kapitel 3** vorstelle, konnten wir umfangreiche Hybridisierung und Introgression in ökologisch vielfältigen Lebensräumen identifizieren. Insgesamt haben wir 49 Genome der drei Arten und deren Hybride von Genotypen aus der Wassersäule und einzelnen Dauereiern aus Sedimentkernen sequenziert und analysiert. Die Introgressionsmuster zeigen eine komplexe Geschichte, die mehrere Generationen der Hybridisierung und Rückkreuzung widerspiegelt.

Basierend auf dieser Arbeit habe ich in **Kapitel 4** die Hybridisierungsdynamik von zwei Populationen ausführlicher untersucht. Ich habe das Referenzgenom weiter verbessert bis auf die Chromosomebene und umfangreichere genomische Daten für eine Zeitserien sequenziert. Ich habe gezeigt, wie eine neue *junctions* Methode, die darauf basiert wie Rekombination

genomische Blöcke verschiedener Herkunft im Laufe der Zeit aufbricht, dafür genutzt werden kann, das Ausmaß und die Zeitpunkte von Hybridisierungsereignissen abzuschätzen. Ich konnte in beiden Populationen wiederholte Hybridisierung über das ganze Genom und über mehrere Jahrzehnte hinweg feststellen sowie vielfältige genomweite Hybrid-Herkunft unter Individuen. Diese Ergebnisse etablieren den *Daphnia longispina* Artenkomplex als ein Hybrid-Modell mit einzigartigen Eigenschaften, wie z.B. die wiederholte Hybridisierung zwischen mehreren Arten in einer Mosaik-Hybridzone, die bisher nicht in Hybridpopulationen beschrieben wurden.

Diese zeitliche Dynamik ist wichtig zu berücksichtigen, da Veränderungen von Hybridzonen durch anthropogene Veränderungen bei *Daphnia* und anderen Hybrid-Modellen beschleunigt werden. Insgesamt zeige ich die zeitliche Dynamik und die vielfältigen Genome von Hybrid-Individuen mit beispielloser Auflösung. Diese Doktorarbeit verdeutlicht die evolutionäre Komplexität, die notwendig ist, um alle Faktoren zu verstehen, die Mosaik-Hybridzonen formen, wie etwa räumliche und zeitliche Veränderungen, wiederholte Hybridisierung und das Vorkommen mehrerer Arten.

Chapter 1 – General introduction

Hybridization and Genomics

Since Charles Darwin's time, biologists have documented hybridization between closely related lineages throughout the tree of life. In animals, this was thought to be both rare and maladaptive. However, with the rise of molecular methods, we have learned about the frequency and importance of hybridization as an evolutionary force (Abbott et al., 2013). Estimates vary, but hybridization and introgression have been reported in approximately 25% of plant species and 10% of animal species (Mallet, 2005). Studying hybridization in different spatial and temporal contexts helps us to understand genetic barriers between taxa and the evolutionary factors, such as selection and recombination, that shape them (Barton & Hewitt, 1985; Schwenk et al., 2008). It can also provide a snapshot of the speciation process, where species may be understood as populations in close geographic proximity with incomplete reproductive barriers that still maintain genetic and phenotypic distinctiveness (Seehausen et al., 2014). Hybridization occurs globally, and some species pairs have diverged for only tens of thousands of years, while others have diverged for millions of years (Harrison & Larson, 2014). This can arise either when populations respond to divergent selection across an environmental gradient or, more commonly, after secondary contact of the parental species following geographical isolation (Durrett et al., 2000). In Europe, for example, many taxa have been shown to hybridize in geographic regions where populations of closely related species that were separated during the last glacial period reestablished gene flow after postglacial range expansion (Hewitt, 2011), e.g., toads (Arntzen et al., 2017) and house mice (Duvaux et al., 2011).

Hybrid zones are generally described as geographic areas where two or more divergent lineages overlap and hybridize, with gradients or clines of genetic composition between the lineages (Barton & Hewitt, 1985; Harrison, 1993). These areas can vary in width from a few meters to hundreds of kilometers or have a patchy distribution of lineages known as mosaic hybrid zones that are often dictated by variations in environmental factors (Harrison, 1986). Replicated hybrid zones, i.e., sampling across multiple hybrid zones or multiple transects in a single hybrid zone, can be particularly useful for better understanding the repeatability of these evolutionary processes (Culumber et al., 2011; Westram et al., 2021).

Outcomes of hybridization

The potential evolutionary outcomes of hybridization are diverse and depend on many factors, such as time since hybridization, population demography, recombination rate, and genome interactions (Figure 1.1). Hybrid genomes in these populations can exhibit varying levels of

ancestry from different species that are shaped by recombination and selection (Runemark et al., 2019).

In some taxa that are able to hybridize, hybridization is rarely observed due to strong pre- or post-zygotic barriers and selection against the hybrids with reduced fitness compared to the parental taxa. Some hybrids can be completely inviable or sterile (Maheshwari & Barbash, 2011), while others display significantly reduced fertility, which can depend on the divergence time between taxa (Dagilis et al., 2019) and hybrid generation (Pritchard et al., 2013). In highly divergent taxa with strong post-zygotic barriers, the accumulation of genetic incompatibilities can also lead to hybrid breakdown after the first generation (F1) (Burton et al., 2013). Nevertheless, simulations using two different models of hybrid incompatibilities predict that species barriers and genomic differentiation can be maintained despite substantial gene flow, depending on the demographic, spatial, and temporal context of hybridization (Lindtke & Buerkle, 2015). In all these cases, the species boundaries are largely maintained, but the species are negatively affected by reduced reproductive success. This can lead to demographic swamping, in which the overall population growth rate is reduced and genetic diversity is lost.

The formation of stable hybrid zones is well documented in plants (reviewed in Abbott, 2017) and animals, including birds, insects, mammals, and reptiles (Brelsford & Irwin, 2009; Harrison & Arnold, 1982; Turner & Harr, 2014; Yang et al., 2020). Different models, such as the tension zone (Barton & Hewitt, 1985) and bounded hybrid superiority model (Moore, 1977), aim to explain how stability in hybrid zones is maintained using different assumptions about selection pressure on hybrids and dispersal of species (Curry, 2015).

After prolonged periods of admixture, the parental species can disappear and the admixed individuals form a hybrid swarm, a population of genetically highly variable later-generation hybrids (Seehausen et al., 2008). Eventually, this can lead to genetic swamping, in which one or both species are replaced by hybrids, which may increase the extinction risk of the parental species (Todesco et al., 2016) or result in speciation reversal (Seehausen et al., 2008).

Hybridization can also result in introgression, the exchange of alleles between taxa via backcrossing following hybridization, in specific genomic regions (Le Moan et al., 2021) or across the entire genome (McFarlane et al., 2021). While these alleles are often neutral, they can also accelerate adaptation and increase fitness for the species when beneficial alleles are exchanged, a phenomenon known as adaptive introgression (Chhatre et al., 2018; Racimo et al., 2017; Valencia-Montoya et al., 2020). Genomic regions that remain highly divergent between hybridizing species indicate that introgression is restricted in these regions and they likely contribute to adaptation and reproductive isolation (Harrison & Larson, 2016).

Ancient hybridization is also thought to play a key role in speciation by providing old genetic variation that can be reassembled into new combinations (Marques et al., 2019), thus facilitating adaptive radiations, for example in fish (De-Kayne et al., 2022; Meier et al., 2017) and birds (Rubin et al., 2022). In some very rare cases, hybridization has given rise to new homoploid hybrid animal species that are reproductively isolated (Mallet, 2007), e.g., Italian sparrows (Elgvin et al., 2017) and Peruvian fur seals (Lopes et al., 2023).

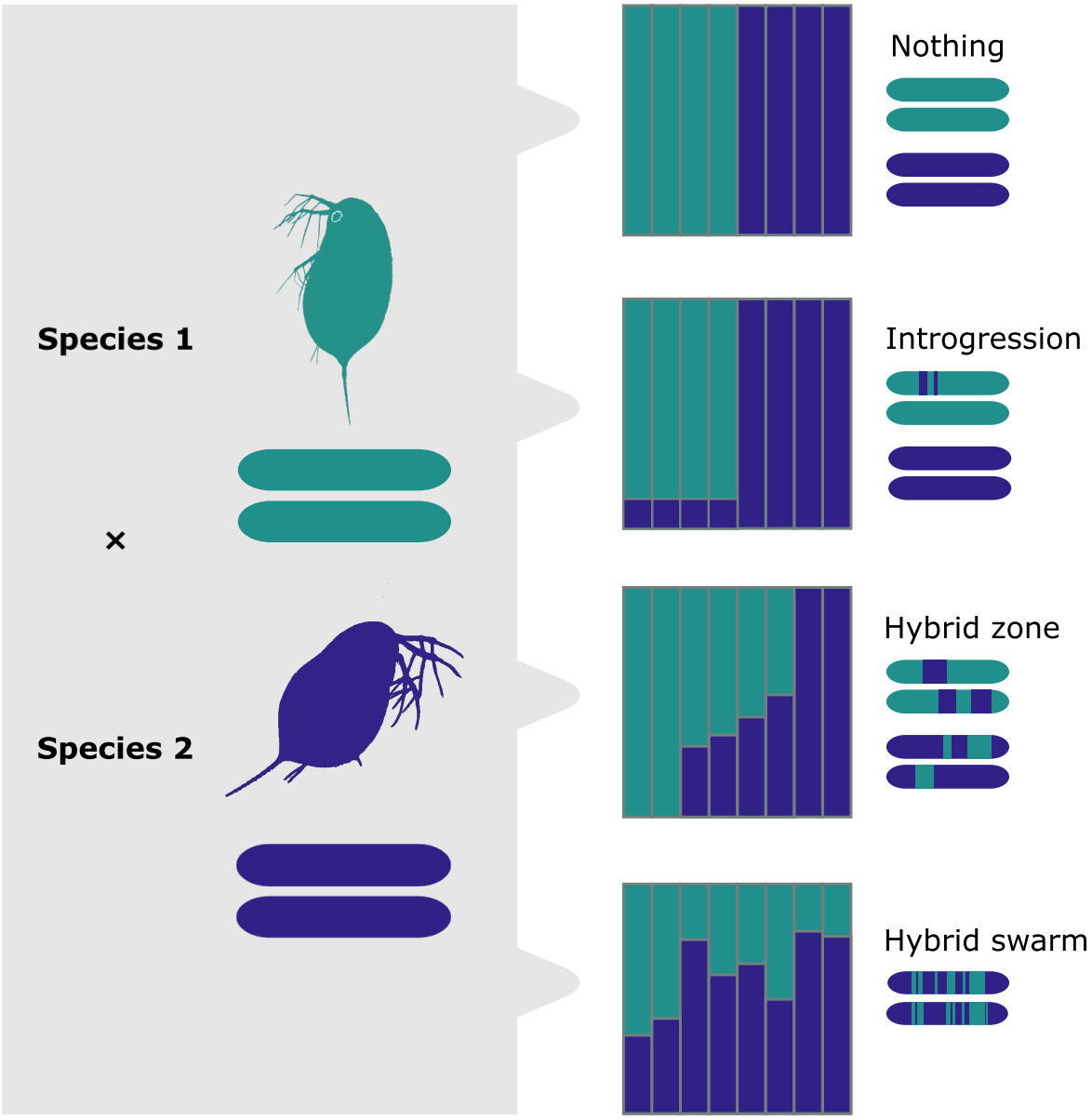


Figure 1.1. Common evolutionary consequences following hybridization. Hybridization is illustrated between two diploid organisms (species 1 and species 2) and a representative chromosome pair is shown. Potential hybridization outcomes are represented by individual admixture proportions for the population and genomic hybridization patterns for a chromosome pair. This figure is inspired by Figure 1 from Runemark et al. (2019).

Genomic tools to study hybridization

To understand the diverse evolutionary processes in hybrid zones, biologists first documented hybridization using morphological characteristics and early molecular tools. With the introduction of high-throughput sequencing in non-model organisms, it became possible to move from single genetic markers to genomic data, such as single nucleotide polymorphism (SNP) and restriction site associated DNA (RAD) markers, to study hybridization frequency, patterns across taxa, and genome composition (Twyford & Ennos, 2012). However, genomic data varies in quality and resolution. Large SNP sets aligned to high-quality annotated reference genomes are currently the most useful for untangling the role of hybridization in the evolutionary history of taxa and the heterogeneous genomic landscapes of admixture from different hybrid zone locations (S. A. Taylor & Larson, 2019), such as those used for honey bees (Calfee et al., 2020) and North American woodpeckers (Aguillon et al., 2021). Many hybridizing species lack reference genomes, and these methods are not yet possible. Recent approaches for generating genome assemblies often involve combining long- and short-read sequence data, allowing for high-quality, contiguous genomes with few errors (Whibley et al., 2021). Chromosome-scale assemblies, which can be achieved using additional scaffolding methods, are especially useful for furthering our understanding of the genomic architecture in hybrids.

Several bioinformatic approaches are available to analyze such genomic data. Each has different requirements and limitations to characterize different aspects of hybridization, such as individual ancestries, direction and timing of gene flow, or ancestry patterns across the genome (Gompert et al., 2017; Payseur & Rieseberg, 2016; Runemark et al., 2019). Most analyses require prior assignment of samples to taxa. Commonly used approaches to detect and quantify hybridization are Bayesian clustering methods such as Structure and ADMIXTURE (Alexander & Lange, 2011; Falush et al., 2003), which estimate the likelihood of individual ancestry based on genetic groups. Hybrid indices can also be used to quantify the proportion of the genome inherited from the parental species (Buerkle, 2005).

Summary statistics are often used to examine patterns of nucleotide diversity (π), genetic differentiation (F_{ST}), and divergence (d_{XY}) between populations, which can indicate gene flow across the genome (Ellegren et al., 2012; J. B. W. Wolf & Ellegren, 2017). Highly differentiated loci or regions can be extracted using F_{ST} outlier analyses, which identify whether the F_{ST} values are larger than expected from genetic drift alone and therefore potentially under selection (Excoffier et al., 2009). However, selection alone is often not the only reason behind differentiation outliers; thus, interpreting these patterns can be challenging because similar genomic landscapes can be shaped by different demographic histories, genomic features, and processes (Ravinet et al., 2017).

To test for introgression across the genome, statistics such as ABBA-BABA (Martin et al., 2015; Patterson et al., 2012) and Patterson's *D* (Green et al., 2010) can be used, which are based on the detection of excess allele sharing between taxa to differentiate introgression from incomplete lineage sorting.

By inferring local ancestry, the resulting ancestry blocks that are broken down by recombination over time can provide insight into processes of genomic admixture and be used to estimate the number of generations since admixture (Corbett-Detig & Nielsen, 2017; Hvala et al., 2018; Pool & Nielsen, 2009). The number of junctions, i.e., breakpoints between ancestry blocks, increases from generation to generation, while the ancestry tract length decreases (Janzen et al., 2018). Recombination varies among species, populations, and individuals (Stapley et al., 2017), and the highly variable recombination rate along genomes drives local patterns of selection and introgression (Martin et al., 2019; Schumer et al., 2018). Local recombination rate variation can be highly species-specific, but some general patterns have been identified. Regions in the center of chromosomes are generally recombination cold spots, while the chromosome peripheries tend to have a higher recombination rate (Haenel et al., 2018). Wersebe et al. (2023) observed that in *Daphnia*, regions of high gene density and recombination rate exhibited higher divergence and differentiation than regions of suppressed recombination. Whether introgressed loci persist in populations or are purged from the genome is therefore highly dependent on their local genomic context. Theory predicts that gene flow between species is significantly reduced if hybrids are less fit and a large number of genomic loci contribute to reproductive barriers (Barton & Bengtsson, 1986). Thus, a broad correlation between introgression and recombination rates can be expected under these conditions, which is indeed found in *Heliconius* butterflies (Martin et al., 2019), house mice (Janoušek et al., 2015), and *Mimulus* monkeyflowers (Brandvain et al., 2014).

Combining different approaches to analyze the genomics of hybrid zones provides a powerful tool for understanding the genomic basis of adaptation.

Anthropogenic change and hybridization

Human activities have significantly shaped the planet's environment and biodiversity, especially since the beginning of the 20th century. The increasing number of environments modified by humans can lead to hybridization of previously isolated species, accelerating the formation of new hybrid zones and impacting existing ones across a range of taxa (Grabenstein & Taylor, 2018; McFarlane & Pemberton, 2019). The most common human-mediated changes that promote hybridization are habitat change and the introduction of non-native species (Ottenburghs, 2021). In contrast to natural hybridization, anthropogenic hybridization often occurs over shorter timescales that can be traced since the initial disturbance. Habitat change

can be direct physical changes, such as deforestation (Sivyer et al., 2018) and urbanization (Päckert et al., 2019), or modifications of the ecological factors, such as climate change (Ryan et al., 2018) and eutrophication (Feulner & Seehausen, 2019).

While introgression is one of the most common outcomes (Ottenburghs, 2021), the diagnostic genetic markers used in many studies are often insufficient to identify backcrossed hybrids (McFarlane & Pemberton, 2019), and high-resolution genomic data is needed to detect hybridization frequency and the number of generations since admixture. In addition, taxa also often have a history of repeated historical admixture, making it challenging to define species for conservation guidelines (Allendorf et al., 2001). Studying both the natural and anthropogenic consequences of ongoing hybridization across taxa can help us unravel the genomic architecture of reproductive isolation.

Hybridization in Freshwater Ecosystems

While freshwater ecosystems hold only a tiny fraction of the total water on Earth, they play a critical role and support approximately 10% of all animal species (Balian et al., 2008). Lakes, ponds, rivers, and wetlands are dynamic ecosystems that fluctuate, e.g., regarding water depth and temperature, and display different levels of habitat complexity and physical connectivity (Grummer et al., 2019). These ecosystems and the species that inhabit them are highly affected by human-mediated changes such as climate change (Woodward et al., 2010) and pollution (Amoatey & Baawain, 2019). In addition, eutrophication, where increased nutrient levels, primarily phosphorus and nitrogen, lead to increased algal and cyanobacterial growth and biomass, is accelerated by agriculture and sewage and can alter the natural trophic state of lakes (Smith, 2003).

Hybridization is common in organisms that spend at least part of their life cycle aquatic, e.g., fish (McKenzie et al., 2015), amphibians (Dufresnes et al., 2021), snails (Westram et al., 2021), and plants (Wu et al., 2022). In contrast to terrestrial hybrid zones, marine hybrid zones are often mosaic with multiple contact zones because animals tend to have much higher dispersal rates, e.g., found in bivalves (Simon et al., 2021) and fish (Riquet et al., 2019). In freshwater ecosystems, hybrid zones can occur along environmental gradients in river systems (Barreto et al., 2020; Culumber et al., 2012), but they are often difficult to identify in the more homogeneous lake habitats with temporally changing environmental conditions (Bittner et al., 2010; Spaak & Hoekstra, 1995).

Like terrestrial hybrid zones, aquatic hybrid zones can be directly affected by anthropogenic changes to the habitat, such as eutrophication, human-built structures, and the deliberate or

accidental introduction of non-native species (Viard et al., 2020). These cases offer the potential to study the outcome of recent hybridization in these essential habitats.

The Model System *Daphnia*

Daphnia (Crustacea, Cladocera) is a genus of small planktonic crustaceans that play an important role in freshwater ecosystems around the world. As key species in the food web, they are primary grazers of phytoplankton and a major food source for planktivorous fish and other invertebrate predators (Lampert & Sommer, 2007). Decades of research into *Daphnia* ecology and evolution as well as their life history traits make them ideally suited as model organisms (Altshuler et al., 2011; Ebert, 2022). The switch between asexual and sexual reproduction allows performing experiments on clonal individuals with the same genetic background, e.g., exposing *Daphnia* to aquatic stressors to explore changes in life history and gene expression, as well as examining population history using resting eggs (Miner et al., 2012). In addition, they exhibit substantial genetic variation (Chaturvedi et al., 2021) and can rapidly adapt to local environmental conditions (De Meester et al., 1999; Wersebe & Weider, 2023).

The genus is comprised of the subgenera *Ctenodaphnia* and *Daphnia*, which diverged around 145 million years ago (Cornetti et al., 2019). Approximately 40 species are found in the subgenus *Daphnia,* which is separated into the *D. pulex* group sensu lato and *D. longispina* group sensu lato (Adamowicz et al., 2009; Chin & Cristescu, 2021).

Most of the year *Daphnia* reproduce asexually through parthenogenesis and switch to the production of males when environmental conditions deteriorate, e.g., overcrowding and changes in photoperiod or food level (Figure 1.2; Decaestecker et al., 2009; Ebert, 2005). Mating results in up to two resting eggs that are encased in a protective chitinous shell called the ephippium, which can remain viable for decades and, in extreme cases, even centuries (Frisch et al., 2014). This results in a greater dispersal ability compared to organisms of similar size through the ephippia, which float on water bodies and are dispersed by water currents, wind, or animals (Slusarczyk et al., 2019). However, a portion of the resting eggs are eventually buried in the lake sediments and contribute to the resting egg bank (Hairston, 1996). If resting eggs from the surface layers or deeper layers after sediment disturbance are exposed to hatching cues, e.g., temperature and changes in light, they can contribute to genetic diversity in the active population and impact population dynamics (Brendonck & De Meester, 2003; Cáceres, 1998; Hairston & Kearns, 2002).

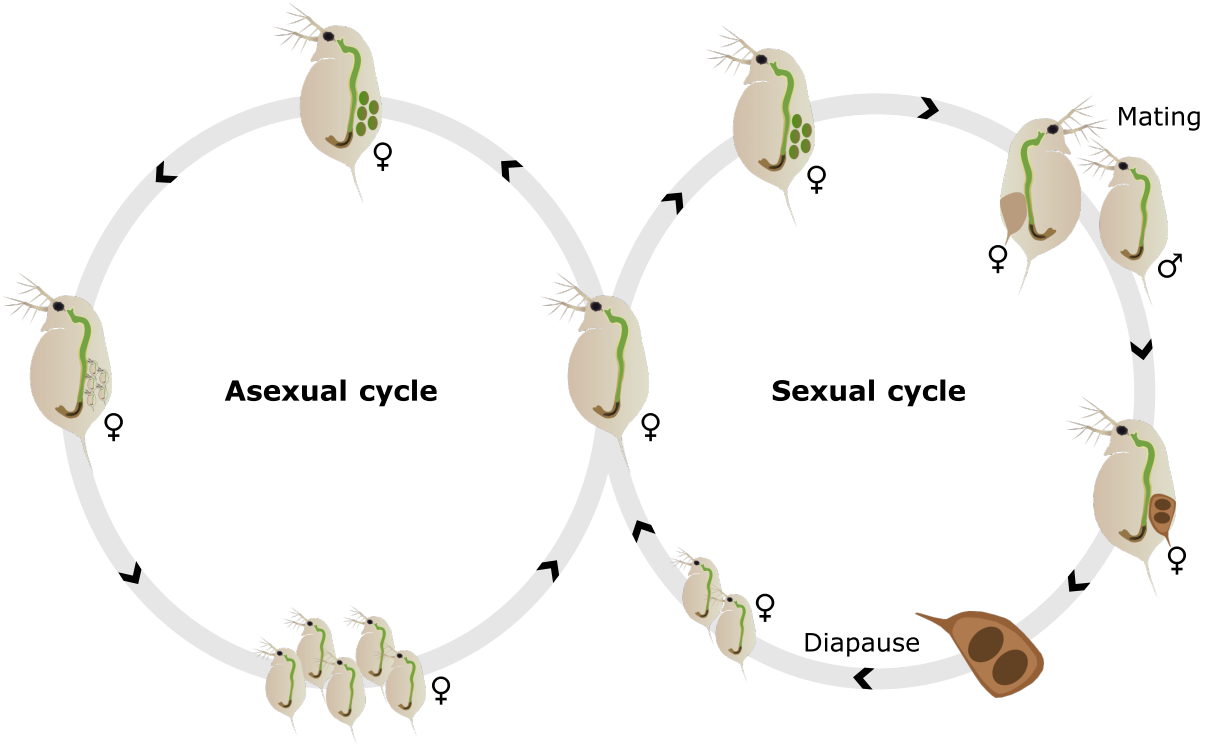


Figure 1.2. The *Daphnia* life cycle. During asexual (clonal) reproduction, adult female *Daphnia* produce parthenogenetic eggs, which develop into embryos and are then released from the brood chamber as parthenogenetic females. Sexual reproduction is triggered when environmental conditions deteriorate and parthenogenetic males are produced. Mating results in fertilized resting eggs encased in the ephippium, from which one or two sexual females will hatch after undergoing diapause.

Many organisms living in aquatic ecosystems are able to produce dormant propagules, such as eggs, seeds, cysts, or spores, which form biological archives and can be used to directly study demographic and adaptive changes in ecosystems (Brendonck & De Meester, 2003; Burge et al., 2018; Orsini et al., 2013). In addition, paleolimnological methods can be used to date sediments and reconstruct biotic and abiotic conditions.

In *Daphnia*, the resting egg bank can be used to hatch historical genotypes from sediment cores to examine evolutionary response, e.g., to increased levels of cyanobacterial food (Isanta-Navarro et al., 2021) and thermal stress (Yousey et al., 2018). Furthermore, DNA can be isolated from resting eggs, and genetic markers, such as microsatellites (Limburg & Weider, 2002; Möst et al., 2015), can be used to investigate genetic variation associated with changes in environmental factors, e.g., cyanobacteria (Monchamp et al., 2017), eutrophication (Alric et al., 2016; Brede et al., 2009), temperature (Dziuba et al., 2020), and invasion of non-native species (Duffy et al., 2000). However, depending on the *Daphnia* species, ephippia measure only about 0.5 to 1.3 mm in length (Tsugeki et al., 2021), and the small amount of total DNA contained in the resting eggs makes genomic methods very challenging without pooling multiple resting eggs (Cordellier et al., 2021). While whole-genome sequencing of single resting eggs using whole genome amplification (WGA) has been shown to work on a small

scale, population genomic investigations are still hindered by contamination and degraded DNA (Lack et al., 2018).

The *Daphnia longispina* Species Complex

Hybridization and introgression are commonly found in the *Daphnia longispina* species complex (DLSC), which inhabits lakes and ponds throughout the Holarctic (Adamowicz et al., 2009; Petrusek, Hobæk, et al., 2008). The complex radiated approximately 5–7 million years ago (Schwenk, 1993) from multiple Palearctic glacial refugia (Petrusek et al., 2012) and consists of the three most frequently found and well-studied species, *D. galeata*, *D. cucullata*, and *D. longispina* (the latter also previously described as *D. hyalina*, *D. rosea*, or *D. zschokkei*), as well as several cryptic and endemic lineages (D. J. Taylor & Hebert, 1994; Zuykova et al., 2019). A comprehensive phylogeny based on mitochondrial markers is presented by Chin and Cristescu (2021).

While the species can co-occur in habitats, they display clear genetic differentiation and differ in ecological preferences. *D. galeata* prefers more eutrophic conditions to establish successfully, while *D. longispina* survives from oligotrophic to eutrophic conditions (Spaak et al., 2012). *D. cucullata* is found in a range of ecological conditions from oligotrophic to hypertrophic lakes (Karpowicz et al., 2020). In addition to trophic level, which is closely linked to food availability, food quality (Brzeziński & Von Elert, 2007; Seidendorf et al., 2010) and the presence of filamentous cyanobacteria (Brzeziński, 2015) also contribute to niche separation. The presence of planktivorous fish shapes the species distribution, with *D. cucullata*, as the smallest of the three species, being less susceptible to predation pressure from visual predators (Spaak & Hoekstra, 1997) and *D. longispina* being the most sensitive to fish predation (Petrusek, Seda, et al., 2008). In contrast, the larger species are less likely to be preyed upon by gap-limited invertebrate predators, such as *Chaoborus* larvae (Pijanowska, 1990; Wolinska et al., 2007). The species also display differences in susceptibility to parasites (Wolinska et al., 2006).

When two or more species appear in the same habitats, they also exhibit further niche separation, e.g., at which time of year the sexual reproduction takes place (Jankowski & Straile, 2004; Macháček et al., 2013) and spatial distribution across horizontal ecological gradients (Petrusek, Seda, et al., 2008) or vertically within the water column (Weider & Stich, 1992).

Hybridization in the DLSC

Hybridization in *Daphnia* has been documented between the North American *D. pulex* species complex, resulting in reticulate evolution (Vergilino et al., 2011) and lineages with polyploidy

and obligate asexual reproduction (Dufresne & Hebert, 1997). It has also been found to occur after anthropogenic disturbance has weakened ecological barriers (Millette et al., 2020).

However, hybridization is most prevalent in the DLSC and has shaped the diversification with introgression between lineages (Ishida & Taylor, 2007) and a history of ancient hybridization (Beninde, 2021). Researchers first documented the occurrence of intermediate morphological phenotypes (Flößner & Kraus, 1986; H. G. Wolf, 1987) and later found widespread evidence for recent hybridization using early genetic markers (Gießler, 1997; Schwenk, 1993; Schwenk et al., 1995; Spaak & Hoekstra, 1995).

Later, nuclear, mitochondrial, and microsatellite markers for the DLSC were developed to distinguish the species and their hybrids (Brede et al., 2006; Rusek et al., 2015; Schwenk et al., 1998; Skage et al., 2007) and investigate species distribution in different habitats (Keller et al., 2008; Petrusek, Seda, et al., 2008; Yin et al., 2010) and genetic structure over time (Griebel et al., 2016; Yin et al., 2014). The resting egg bank also allowed the investigation of species composition and how hybridization has changed over the past decades (Brede et al., 2009; Cousyn et al., 2001). Nevertheless, these methods are only able to reliably assign individuals as F1, F2, or backcrosses (Dlouhá et al., 2010) and are not sensitive enough to identify genome-wide signatures of introgression (Twyford & Ennos, 2012).

In addition to ecological barriers, the species are isolated by reproductive barriers, with differences in viability depending on the divergence time between species pairs and cross direction (Chin & Cristescu, 2021). *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata* hybrids are less likely to produce resting eggs, hatch, and survive compared to the parental species in the field (Keller et al., 2007; Spaak et al., 2004) and the laboratory (Schwenk et al., 2001). *D. longispina* × *D. cucullata* hybrids are rarely documented in natural populations, most likely due to their opposing ecological preferences (Petrusek, Seda, et al., 2008). It is not clear to what extent introgression occurs between all possible species pairs and how it varies across the genome.

We find a patchy distribution of freshwater ecosystems with different environmental conditions that are colonized by *Daphnia* at different rates, followed by selection leading to mosaic hybrid zones (Harrison, 1986). These hybrids often show intermediate morphological and ecological traits and are found in intermediate habitats (Löffler et al., 2004; Ma et al., 2019). This could be explained by the tension zone model, which posits that hybrid zones are maintained by random dispersal of the parental species and selection against the less fit hybrid individuals (Barton & Hewitt, 1985). However, due to their clonal reproduction phase, hybrids can temporarily outperform their parental species and dominate lake habitats under certain environmental conditions (Declerck & De Meester, 2003; Griebel et al., 2015). Furthermore, in

some habitats, these hybridization dynamics have been influenced by increased levels of anthropogenic eutrophication in the 20th century (Alric et al., 2016; Keller et al., 2008).

The small genome size of *Daphnia* (~190 Mb for *D. pulex*, Ye et al., 2017) and decreasing sequencing costs make population genomic studies more feasible and affordable. Since the publication of the *D. pulex* genome (Colbourne et al., 2011), the first crustacean genome to be sequenced, additional genomic resources have been created to advance *Daphnia* as a genomic model organism. These include the *D. magna* (Lee et al., 2019) and *D. pulicaria* genomes (Wersebe et al., 2023), as well as the *D. galeata* transcriptome (Huylmans et al., 2016). In *D. pulex*, studies have investigated genome-wide patterns of differentiation using whole-genome sequencing (Maruki et al., 2022) and ddRAD-seq (Wersebe et al., 2023). However, a reference genome is needed to investigate genome-wide hybridization and introgression in the DLSC.

Thesis Outline and Objective

The aim of this thesis is to establish the *Daphnia longispina* species complex as a genomic model for hybridization by providing a method for sequencing resting eggs and a chromosome-scale genome assembly for the group. This serves as the basis for understanding the extensive history of hybridization in the DLSC and the genomic patterns of recent hybridization and introgression in mosaic hybrid zones.

In **Chapter 2**, I developed a novel method to address issues with low DNA yield and degradation and generate whole genomes from single DLSC resting eggs using whole genome amplification. Using a pre-sequencing contamination screening, a cost-saving protocol was established to optimize the population genomic investigation of the resting egg bank in *Daphnia* and thus the ability to track past hybridization events.

In **Chapter 3**, I present a high-quality genome assembly and annotation for the species *D. galeata*, the first for a member of the DLSC. Using the protocol established in **Chapter 2** and the new reference genome, I generated and analyzed whole genomes of three species of the complex to detect hybridization frequency and investigate introgression and divergence patterns across the genome in four populations.

After identifying two populations with high hybridization frequency in **Chapter 3**, I conducted more extensive time-series genomic sampling of these populations for **Chapter 4** and improved the reference genome to a chromosome scale. By refining a novel junctions method, I aim to accurately date recent admixture events to understand the history of hybridization and the distribution of ancestry across the genome.

Introduction References

Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, *55*(4), 238–258. https://doi.org/10.1111/jse.12267

Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Adamowicz, S. J., Petrusek, A., Colbourne, J. K., Hebert, P. D. N., & Witt, J. D. S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, *50*(3), 423–436. https://doi.org/10.1016/j.ympev.2008.11.026

Aguillon, S. M., Walsh, J., & Lovette, I. J. (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1943), 20201805. https://doi.org/10.1098/rspb.2020.1805

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Alric, B., Möst, M., Domaizon, I., Pignol, C., Spaak, P., & Perga, M.-E. (2016). Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology*, *29*(4), 720–735. https://doi.org/10.1111/jeb.12820

Altshuler, I., Demiri, B., Xu, S., Constantin, A., Yan, N. D., & Cristescu, M. E. (2011). An Integrated Multi-Disciplinary Approach for Studying Multiple Stressors in Freshwater Ecosystems: *Daphnia* as a Model Organism. *Integrative and Comparative Biology*, *51*(4), 623–633. https://doi.org/10.1093/icb/icr103

Amoatey, P., & Baawain, M. S. (2019). Effects of pollution on freshwater aquatic organisms. *Water Environment Research*, *91*(10), 1272–1287. https://doi.org/10.1002/wer.1221

Arntzen, J. W., De Vries, W., Canestrelli, D., & Martínez-Solano, I. (2017). Hybrid zone formation and contrasting outcomes of secondary contact over transects in common toads. *Molecular Ecology*, *26*(20), 5663–5675. https://doi.org/10.1111/mec.14273

Balian, E. V., Segers, H., Lévèque, C., & Martens, K. (2008). The Freshwater Animal Diversity Assessment: An overview of the results. *Hydrobiologia*, *595*(1), 627–637. https://doi.org/10.1007/s10750-007-9246-3

Barreto, S. B., Knowles, L. L., Affonso, P. R. A. D. M., & Batalha-Filho, H. (2020). Riverscape properties contribute to the origin and structure of a hybrid zone in a Neotropical freshwater fish. *Journal of Evolutionary Biology*, *33*(11), 1530–1542. https://doi.org/10.1111/jeb.13689

Barton, N. H., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(3), 357–376. https://doi.org/10.1038/hdy.1986.135

Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Beninde, J. (2021). *Evolution of the Daphnia longispina Group in Response to Anthropogenic Environmental Changes* [Ph.D. thesis, Universität Konstanz]. https://kops.uni-konstanz.de/handle/123456789/54238

Bittner, D., Excoffier, L., & Largiadèr, C. R. (2010). Patterns of morphological changes and hybridization between sympatric whitefish morphs (*Coregonus* spp.) in a Swiss lake: A role for eutrophication? *Molecular Ecology*, *19*(10), 2152–2167. https://doi.org/10.1111/j.1365-294X.2010.04623.x

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics*, *10*(6), e1004410. https://doi.org/10.1371/journal.pgen.1004410

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences*, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Brede, N., Thielsch, A., Sandrock, C., Spaak, P., Keller, B., Streit, B., & Schwenk, K. (2006). Microsatellite markers for European *Daphnia*. *Molecular Ecology Notes*, *6*(2), 536–539. https://doi.org/10.1111/j.1471-8286.2005.01218.x

Brelsford, A., & Irwin, D. E. (2009). Incipient Speciation despite Little Assortative Mating: The Yellow-Rumped Warbler Hybrid Zone. *Evolution*, *63*(12), 3050–3060.

Brendonck, L., & De Meester, L. (2003). Egg banks in freshwater zooplankton: Evolutionary and ecological archives in the sediment. *Hydrobiologia*, *491*(1), 65–84. https://doi.org/10.1023/A:1024454905119

Brzeziński, T. (2015). Filamentous cyanobacteria alter the relative fitness in a *Daphnia* hybrid species complex. *Freshwater Biology*, *60*(1), 101–110. https://doi.org/10.1111/fwb.12471

Brzeziński, T., & Von Elert, E. (2007). Biochemical food quality effects on a *Daphnia* hybrid complex. *Limnology and Oceanography*, *52*(6), 2350–2357. https://doi.org/10.4319/lo.2007.52.6.2350

Buerkle, C. A. (2005). Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, *5*(3), 684–687. https://doi.org/10.1111/j.1471-8286.2005.01011.x

Burge, D. R. L., Edlund, M. B., & Frisch, D. (2018). Paleolimnology and resurrection ecology: The future of reconstructing the past. *Evolutionary Applications*, *11*(1), 42–59. https://doi.org/10.1111/eva.12556

Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 281–302. https://doi.org/10.1146/annurev-ecolsys-110512-135758

Cáceres, C. E. (1998). Interspecific Variation in the Abundance, Production, and Emergence of *Daphnia* Diapausing Eggs. *Ecology*, *79*(5), 1699–1710. https://doi.org/10.1890/0012-9658(1998)079[1699:IVITAP]2.0.CO;2

Calfee, E., Agra, M. N., Palacio, M. A., Ramírez, S. R., & Coop, G. (2020). Selection and hybridization shaped the rapid spread of African honey bee ancestry in the Americas. *PLoS Genetics*, *16*(10), e1009038. https://doi.org/10.1371/journal.pgen.1009038

Chaturvedi, A., Zhou, J., Raeymaekers, J. A. M., Czypionka, T., Orsini, L., Jackson, C. E., Spanier, K. I., Shaw, J. R., Colbourne, J. K., & De Meester, L. (2021). Extensive standing genetic variation from a small number of founders enables rapid adaptation in *Daphnia*. *Nature Communications*, *12*(1), 4306. https://doi.org/10.1038/s41467-021-24581-z

Chhatre, V. E., Evans, L. M., DiFazio, S. P., & Keller, S. R. (2018). Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Molecular Ecology*, *27*(23), 4820–4838. https://doi.org/10.1111/mec.14820

Chin, T. A., & Cristescu, M. E. (2021). Speciation in *Daphnia*. *Molecular Ecology*, *30*(6), 1398–1418. https://doi.org/10.1111/mec.15824

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cáceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., … Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, *331*(6017), 555–561. https://doi.org/10.1126/science.1197761

Corbett-Detig, R., & Nielsen, R. (2017). A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genetics*, *13*(1), e1006529. https://doi.org/10.1371/journal.pgen.1006529

Cordellier, M., Wojewodzic, M. W., Wessels, M., Kuster, C., & Von Elert, E. (2021). Next-generation sequencing of DNA from resting eggs: Signatures of eutrophication in a lake's sediment. *Zoology*, *145*, 125895. https://doi.org/10.1016/j.zool.2021.125895

Cornetti, L., Fields, P. D., Van Damme, K., & Ebert, D. (2019). A fossil-calibrated phylogenomic analysis of *Daphnia* and the Daphniidae. *Molecular Phylogenetics and Evolution*, *137*, 250–262. https://doi.org/10.1016/j.ympev.2019.05.018

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D., & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, *98*(11), 6256–6260. https://doi.org/10.1073/pnas.111606798

Culumber, Z. W., Fisher, H. S., Tobler, M., Mateos, M., Barber, P. H., Sorenson, M. D., & Rosenthal, G. G. (2011). Replicated hybrid zones of *Xiphophorus* swordtails along an elevational gradient. *Molecular Ecology*, *20*(2), 342–356. https://doi.org/10.1111/j.1365-294X.2010.04949.x

Culumber, Z. W., Shepard, D. B., Coleman, S. W., Rosenthal, G. G., & Tobler, M. (2012). Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: *Xiphophorus*). *Journal of Evolutionary Biology*, *25*(9), 1800–1814. https://doi.org/10.1111/j.1420-9101.2012.02562.x

Curry, C. M. (2015). An Integrated Framework for Hybrid Zone Models. *Evolutionary Biology*, *42*(3), 359–365. https://doi.org/10.1007/s11692-015-9332-9

Dagilis, A. J., Kirkpatrick, M., & Bolnick, D. I. (2019). The evolution of hybrid fitness during speciation. *PLoS Genetics*, *15*(5), e1008125. https://doi.org/10.1371/journal.pgen.1008125

De Meester, L., Boersma, M., & Spaak, P. (1999). Environmental stress and local adaptation in *Daphnia magna*. *Limnology and Oceanography*, *44*(2), 393–402. https://doi.org/10.4319/lo.1999.44.2.0393

Decaestecker, E., De Meester, L., & Mergeay, J. (2009). Cyclical Parthenogenesis in *Daphnia*: Sexual Versus Asexual Reproduction. In I. Schön, K. Martens, & P. Dijk (Eds.), *Lost Sex* (pp. 295–316). Springer Netherlands. https://doi.org/10.1007/978-90-481-2770-2_15

Declerck, S., & De Meester, L. (2003). Impact of fish predation on coexisting *Daphnia* taxa: A partial test of the temporal hybrid superiority hypothesis. *Hydrobiologia*, *500*(1), 83–94. https://doi.org/10.1023/A:1024656602248

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications*, *13*(1), 4479. https://doi.org/10.1038/s41467-022-32181-8

Dlouhá, Š., Thielsch, A., Kraus, R. H. S., Seda, J., Schwenk, K., & Petrusek, A. (2010). Identifying hybridizing taxa within the *Daphnia longispina* species complex: A comparison of genetic methods and phenotypic approaches. *Hydrobiologia*, *643*(1), 107–122. https://doi.org/10.1007/s10750-010-0128-8

Duffy, M. A., Perry, L. J., Kearns, C. M., Weider, L. J., & Hairston, N. G. (2000). Paleogenetic evidence for a past invasion of Onondaga Lake, New York, by exotic *Daphnia curvirostris* using mtDNA from dormant eggs. *Limnology and Oceanography*, *45*(6), 1409–1414. https://doi.org/10.4319/lo.2000.45.6.1409

Dufresne, F., & Hebert, P. D. N. (1997). Hybridization and origins of polyploidy. *Proceedings of the Royal Society of London B: Biological Sciences*, *258*(1352), 141–146. https://doi.org/10.1098/rspb.1994.0154

Dufresnes, C., Suchan, T., Smirnov, N. A., Denoël, M., Rosanov, J. M., & Litvinchuk, S. N. (2021). Revisiting a speciation classic: Comparative analyses support sharp but leaky transitions between *Bombina* toads. *Journal of Biogeography*, *48*(3), 548–560. https://doi.org/10.1111/jbi.14018

Durrett, R., Buttel, L., & Harrison, R. (2000). Spatial models for hybrid zones. *Heredity*, *84*(1), 9–19. https://doi.org/10.1046/j.1365-2540.2000.00566.x

Duvaux, L., Belkhir, K., Boulesteix, M., & Boursot, P. (2011). Isolation and gene flow: Inferring the speciation history of European house mice. *Molecular Ecology*, *20*(24), 5248–5264. https://doi.org/10.1111/j.1365-294X.2011.05343.x

Dziuba, M. K., Herdegen-Radwan, M., Pluta, E., Wejnerowski, Ł., Szczuciński, W., & Cerbin, S. (2020). Temperature increase altered *Daphnia* community structure in artificially heated lakes: A potential scenario for a warmer future. *Scientific Reports*, *10*(1), 13956. https://doi.org/10.1038/s41598-020-70294-6

Ebert, D. (2005). *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*. National Library of Medicine (US), National Center for Biotechnology. https://www.ncbi.nlm.nih.gov/books/NBK2036/

Ebert, D. (2022). *Daphnia* as a versatile model system in ecology and evolution. *EvoDevo*, *13*(1), 16. https://doi.org/10.1186/s13227-022-00199-0

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*(7426), 756–760. https://doi.org/10.1038/nature11584

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285–298. https://doi.org/10.1038/hdy.2009.74

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, *164*(4), 1567–1587. https://doi.org/10.1093/genetics/164.4.1567

Feulner, P. G. D., & Seehausen, O. (2019). Genomic insights into the vulnerability of sympatric whitefish species flocks. *Molecular Ecology*, *28*(3), 615–629. https://doi.org/10.1111/mec.14977

Flößner, D., & Kraus, K. (1986). On the taxonomy of the *Daphnia hyalina-galeata* complex (Crustacea: Cladocera). *Hydrobiologia*, *137*(2), 97–115. https://doi.org/10.1007/BF00004206

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368. https://doi.org/10.1111/ele.12237

Gießler, S. (1997). Analysis of reticulate relationships within the *Daphnia longispina* species complex. Allozyme phenotype and morphology. *Journal of Evolutionary Biology*, *10*(1), 87–105. https://doi.org/10.1046/j.1420-9101.1997.10010087.x

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking Barriers: Causes, Consequences, and Experimental Utility of Human-Mediated Hybridization. *Trends in Ecology & Evolution*, *33*(3), 198–212. https://doi.org/10.1016/j.tree.2017.12.008

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Griebel, J., Gießler, S., Poxleitner, M., Faria, A. N., Yin, M., & Wolinska, J. (2015). Extreme Environments Facilitate Hybrid Superiority – The Story of a Successful *Daphnia galeata* × *longispina* Hybrid Clone. *PLoS One*, *10*(10), e0140275. https://doi.org/10.1371/journal.pone.0140275

Griebel, J., Gießler, S., Yin, M., & Wolinska, J. (2016). Parental and hybrid *Daphnia* from the *D. longispina* complex: Long-term dynamics in genetic structure and significance of overwintering modes. *Journal of Evolutionary Biology*, *29*(4), 810–823. https://doi.org/10.1111/jeb.12828

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. *Trends in Ecology & Evolution*, *34*(7), 641–654. https://doi.org/10.1016/j.tree.2019.02.013

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, *27*(11), 2477–2497. https://doi.org/10.1111/mec.14699

Hairston, N. G. (1996). Zooplankton egg banks as biotic reservoirs in changing environments. *Limnology and Oceanography*, *41*(5), 1087–1092.

Hairston, N. G., & Kearns, C. M. (2002). Temporal Dispersal: Ecological and Evolutionary Aspects of Zooplankton Egg Banks and the Role of Sediment Mixing. *Integrative and Comparative Biology*, *42*(3), 481–491. https://doi.org/10.1093/icb/42.3.481

Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, *56*(3), 337–349. https://doi.org/10.1038/hdy.1986.55

Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process*. Oxford University Press.

Harrison, R. G., & Arnold, J. (1982). A Narrow Hybrid Zone Between Closely Related Cricket Species. *Evolution*, *36*(3), 535–552. https://doi.org/10.2307/2408099

Harrison, R. G., & Larson, E. L. (2014). Hybridization, Introgression, and the Nature of Species Boundaries. *Journal of Heredity*, *105*(S1), 795–809. https://doi.org/10.1093/jhered/esu033

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466. https://doi.org/10.1111/mec.13582

Hewitt, G. M. (2011). Quaternary phylogeography: The roots of hybrid zones. *Genetica*, *139*(5), 617–638. https://doi.org/10.1007/s10709-011-9547-3

Huylmans, A. K., López Ezquerra, A., Parsch, J., & Cordellier, M. (2016). De Novo Transcriptome Assembly and Sex-Biased Gene Expression in the Cyclical Parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution*, *8*(10), 3120–3139. https://doi.org/10.1093/gbe/evw221

Hvala, J. A., Frayer, M. E., & Payseur, B. A. (2018). Signatures of hybridization and speciation in genomic patterns of ancestry. *Evolution*, *72*(8), 1540–1552. https://doi.org/10.1111/evo.13509

Isanta-Navarro, J., Hairston, N. G., Beninde, J., Meyer, A., Straile, D., Möst, M., & Martin-Creuzburg, D. (2021). Reversed evolution of grazer resistance to cyanobacteria. *Nature Communications*, *12*(1), 1945. https://doi.org/10.1038/s41467-021-22226-9

Ishida, S., & Taylor, D. J. (2007). Quaternary diversification in a sexual Holarctic zooplankter, *Daphnia galeata*. *Molecular Ecology*, *16*(3), 569–582. https://doi.org/10.1111/j.1365-294X.2006.03160.x

Jankowski, T., & Straile, D. (2004). Allochronic differentiation among *Daphnia* species, hybrids and backcrosses: The importance of sexual reproduction for population dynamics and genetic architecture. *Journal of Evolutionary Biology*, *17*(2), 312–321.

Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional Organization of the Genome May Shape the Species Boundary in the House Mouse. *Molecular Biology and Evolution*, *32*(5), 1208–1220. https://doi.org/10.1093/molbev/msv011

Janzen, T., Nolte, A. W., & Traulsen, A. (2018). The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. *Evolution*, *72*(4), 735–750. https://doi.org/10.1111/evo.13436

Karpowicz, M., Sługocki, Ł., Kozłowska, J., Ochocka, A., & López, C. (2020). Body size of *Daphnia cucullata* as an indicator of the ecological status of temperate lakes. *Ecological Indicators*, *117*, 106585. https://doi.org/10.1016/j.ecolind.2020.106585

Keller, B., Wolinska, J., Manca, M., & Spaak, P. (2008). Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2943–2952. https://doi.org/10.1098/rstb.2008.0044

Keller, B., Wolinska, J., Tellenbach, C., & Spaak, P. (2007). Reproductive isolation keeps hybridizing *Daphnia* species distinct. *Limnology and Oceanography*, *52*(3), 984–991. https://doi.org/10.4319/lo.2007.52.3.0984

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2018). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, *18*(1), 118–127. https://doi.org/10.1111/1755-0998.12720

Lampert, W., & Sommer, U. (2007). *Limnoecology: The Ecology of Lakes and Streams* (Second Edition). Oxford University Press.

Le Moan, A., Roby, C., Fraïsse, C., Daguin-Thiébaut, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, *30*(24), 6718–6732. https://doi.org/10.1111/mec.16189

Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, *210*, 69–84. https://doi.org/10.1016/j.aquatox.2019.02.009

Limburg, P. A., & Weider, L. J. (2002). 'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1488), 281–287. https://doi.org/10.1098/rspb.2001.1868

Lindtke, D., & Buerkle, C. A. (2015). The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, *69*(8), 1987–2004. https://doi.org/10.1111/evo.12725

Löffler, A., Wolinska, J., Keller, B., Rothhaupt, K.-O., & Spaak, P. (2004). Life history patterns of parental and hybrid *Daphnia* differ between lakes. *Freshwater Biology*, *49*(10), 1372–1380. https://doi.org/10.1111/j.1365-2427.2004.01272.x

Lopes, F., Oliveira, L. R., Beux, Y., Kessler, A., Cárdenas-Alayza, S., Majluf, P., Páez-Rosas, D., Chaves, J., Crespo, E., Brownell, R. L., Baylis, A. M. M., Sepúlveda, M., Franco-Trecu, V., Loch, C., Robertson, B. C., Peart, C. R., Wolf, J. B. W., & Bonatto, S. L. (2023). Genomic evidence for homoploid hybrid speciation in a marine mammal apex predator. *Science Advances*, *9*(18), eadf6601. https://doi.org/10.1126/sciadv.adf6601

Ma, X., Hu, W., Smilauer, P., Yin, M., & Wolinska, J. (2019). *Daphnia galeata* and *D. dentifera* are geographically and ecologically separated whereas their hybrids occur in intermediate habitats: A survey of 44 Chinese lakes. *Molecular Ecology*, *28*(4), 785–802. https://doi.org/10.1111/mec.14991

Macháček, J., Vaníčková, I., Seda, J., Cordellier, M., & Schwenk, K. (2013). Sexual reproduction of *Daphnia* in a deep temperate reservoir: The phenology and genetics of male formation. *Hydrobiologia*, *715*(1), 113–123. https://doi.org/10.1007/s10750-012-1419-z

Maheshwari, S., & Barbash, D. A. (2011). The genetics of hybrid incompatibilities. *Annual Review of Genetics*, *45*, 331–355. https://doi.org/10.1146/annurev-genet-110410-132514

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283. https://doi.org/10.1038/nature05706

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, *17*(2), e2006288. https://doi.org/10.1371/journal.pbio.2006288

Maruki, T., Ye, Z., & Lynch, M. (2022). Evolutionary Genomics of a Subdivided Species. *Molecular Biology and Evolution*, *39*(8), msac152. https://doi.org/10.1093/molbev/msac152

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

McFarlane, S. E., Senn, H. V., Smith, S. L., & Pemberton, J. M. (2021). Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, *30*(9), 2104–2115. https://doi.org/10.1111/mec.15862

McKenzie, J. L., Dhillon, R. S., & Schulte, P. M. (2015). Evidence for a bimodal distribution of hybrid indices in a hybrid zone with high admixture. *Royal Society Open Science*, *2*(12), 150285. https://doi.org/10.1098/rsos.150285

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Millette, K. L., Gonzalez, A., & Cristescu, M. E. (2020). Breaking ecological barriers: Anthropogenic disturbance leads to habitat transitions, hybridization, and high genetic diversity. *Science of The Total Environment*, *740*, 140046. https://doi.org/10.1016/j.scitotenv.2020.140046

Miner, B. E., De Meester, L., Pfrender, M. E., Lampert, W., & Hairston, N. G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1735), 1873–1882. https://doi.org/10.1098/rspb.2011.2404

Monchamp, M.-E., Enache, I., Turko, P., Pomati, F., Rîşnoveanu, G., & Spaak, P. (2017). Sedimentary and egg-bank DNA from 3 European lakes reveal concurrent changes in the composition and diversity of cyanobacterial and *Daphnia* communities. *Hydrobiologia*, *800*(1), 155–172. https://doi.org/10.1007/s10750-017-3247-7

Moore, W. S. (1977). An Evaluation of Narrow Hybrid Zones in Vertebrates. *The Quarterly Review of Biology*, *52*(3), 263–277. https://doi.org/10.1086/409995

Möst, M., Oexle, S., Marková, S., Aidukaite, D., Baumgartner, L., Stich, H.-B., Wessels, M., Martin-Creuzburg, D., & Spaak, P. (2015). Population genetic dynamics of an invasion reconstructed from the sediment egg bank. *Molecular Ecology*, *24*(16), 4074–4093. https://doi.org/10.1111/mec.13298

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J. K., Pfrender, M. E., & Weider, L. J. (2013). The evolutionary time machine: Using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution*, *28*(5), 274–282. https://doi.org/10.1016/j.tree.2013.01.009

Ottenburghs, J. (2021). The genic view of hybridization in the Anthropocene. *Evolutionary Applications*, *14*(10), 2342–2360. https://doi.org/10.1111/eva.13223

Päckert, M., Ait Belkacem, A., Wolfgramm, H., Gast, O., Canal, D., Giacalone, G., Lo Valvo, M., Vamberger, M., Wink, M., Martens, J., & Stuckas, H. (2019). Genetic admixture despite ecological segregation in a North African sparrow hybrid zone (Aves, Passeriformes, *Passer domesticus* × *Passer hispaniolensis*). *Ecology and Evolution*, *9*(22), 12710–12726. https://doi.org/10.1002/ece3.5744

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Petrusek, A., Hobæk, A., Nilssen, J. P., Skage, M., Černý, M., Brede, N., & Schwenk, K. (2008). A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta*, *37*(5), 507–519. https://doi.org/10.1111/j.1463-6409.2008.00336.x

Petrusek, A., Seda, J., Macháček, J., Ruthová, Š., & Šmilauer, P. (2008). *Daphnia* hybridization along ecological gradients in pelagic environments: The potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2931–2941. https://doi.org/10.1098/rstb.2008.0026

Petrusek, A., Thielsch, A., & Schwenk, K. (2012). Mitochondrial sequence variation suggests extensive cryptic diversity within the Western Palearctic *Daphnia longispina* complex. *Limnology and Oceanography*, *57*(6), 1838–1845. https://doi.org/10.4319/lo.2012.57.6.1838

Pijanowska, J. (1990). Cyclomorphosis in *Daphnia*: An adaptation to avoid invertebrate predation. *Hydrobiologia*, *198*(1), 41–50. https://doi.org/10.1007/BF00048621

Pool, J. E., & Nielsen, R. (2009). Inference of Historical Changes in Migration Rate From the Lengths of Migrant Tracts. *Genetics*, *181*(2), 711–719. https://doi.org/10.1534/genetics.108.098095

Pritchard, V. L., Knutson, V. L., Lee, M., Zieba, J., & Edmands, S. (2013). Fitness and morphological outcomes of many generations of hybridization in the copepod *Tigriopus californicus*. *Journal of Evolutionary Biology*, *26*(2), 416–433. https://doi.org/10.1111/jeb.12060

Racimo, F., Marnetto, D., & Huerta-Sánchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Molecular Biology and Evolution*, *34*(2), 296–317. https://doi.org/10.1093/molbev/msw216

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Béduneau, V., Briard, O., El Ayari, T., Hochscheid, S., Belkhir, K., Arnaud-Haond, S., Gagnaire, P.-A., & Bierne, N. (2019). Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. *Evolution*, *73*(4), 817–835. https://doi.org/10.1111/evo.13696

Rubin, C.-J., Enbody, E. D., Dobreva, M. P., Abzhanov, A., Davis, B. W., Lamichhaney, S., Pettersson, M., Sendell-Price, A. T., Sprehn, C. G., Valle, C. A., Vasco, K., Wallerman, O., Grant, B. R., Grant, P. R., & Andersson, L. (2022). Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules. *Science Advances*, *8*(27), eabm5982. https://doi.org/10.1126/sciadv.abm5982

Runemark, A., Vallejo-Marin, M., & Meier, J. I. (2019). Eukaryote hybrid genomes. *PLoS Genetics*, *15*(11), e1008404. https://doi.org/10.1371/journal.pgen.1008404

Rusek, J., Ayan, G. B., Turko, P., Tellenbach, C., Giessler, S., Spaak, P., & Wolinska, J. (2015). New possibilities arise for studies of hybridization: SNP-based markers for the multi-species *Daphnia longispina* complex derived from transcriptome data. *Journal of Plankton Research*, *37*(3), 626–635. https://doi.org/10.1093/plankt/fbv028

Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences*, *115*(10), E2284–E2291. https://doi.org/10.1073/pnas.1714950115

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

Schwenk, K. (1993). Interspecific hybridization in *Daphnia*: Distinction and origin of hybrid matrilines. *Molecular Biology and Evolution*, *10*(6), 1289–1302. https://doi.org/10.1093/oxfordjournals.molbev.a040076

Schwenk, K., Bijl, M., & Menken, S. B. J. (2001). Experimental interspecific hybridization in *Daphnia*. *Hydrobiologia*, *442*(1), 67–73. https://doi.org/10.1023/A:1017594325506

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2805–2811. https://doi.org/10.1098/rstb.2008.0055

Schwenk, K., Ender, A., & Streit, B. (1995). What can molecular markers tell us about the evolutionary history of *Daphnia* species complexes? *Hydrobiologia*, *307*(1), 1–7. https://doi.org/10.1007/BF00031991

Schwenk, K., Sand, A., Boersma, M., Brehm, M., Mader, E., Offerhaus, D., & Spaak, P. (1998). Genetic markers, genealogies and biogeographic patterns in the cladocera. *Aquatic Ecology*, *32*(1), 37–51. https://doi.org/10.1023/A:1009939901198

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, *17*(1), 30–44. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Seidendorf, B., Meier, N., Petrusek, A., Boersma, M., Streit, B., & Schwenk, K. (2010). Sensitivity of *Daphnia* species to phosphorus-deficient diets. *Oecologia*, *162*(2), 349–357. https://doi.org/10.1007/s00442-009-1452-4

Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, *34*(1), 208–223. https://doi.org/10.1111/jeb.13709

Sivyer, L., Morgan-Richards, M., Koot, E., & Trewick, S. A. (2018). Anthropogenic cause of range shifts and gene flow between two grasshopper species revealed by environmental modelling, geometric morphometrics and population genetics. *Insect Conservation and Diversity*, *11*(5), 415–434. https://doi.org/10.1111/icad.12289

Skage, M., Hobæk, A., Ruthová, Š., Keller, B., Petrusek, A., Seda, J., & Spaak, P. (2007). Intra-specific rDNA-ITS restriction site variation and an improved protocol to distinguish species and hybrids in the *Daphnia longispina* complex. *Hydrobiologia*, *594*(1), 19–32. https://doi.org/10.1007/s10750-007-9090-5

Slusarczyk, M., Pinel-Alloul, B., & Pietrzak, B. (2019). Mechanisms Facilitating Dispersal of Dormant Eggs in a Planktonic Crustacean. In V. R. Alekseev & B. Pinel-Alloul (Eds.), *Dormancy in Aquatic Organisms. Theory, Human Use and Modeling* (pp. 137–161). Springer. https://doi.org/10.1007/978-3-030-21213-1_8

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems: A global problem. *Environmental Science and Pollution Research International*, *10*(2), 126–139. https://doi.org/10.1065/espr2002.12.142

Spaak, P., Denk, A., Boersma, M., & Weider, L. J. (2004). Spatial and temporal patterns of sexual reproduction in a hybrid *Daphnia* species complex. *Journal of Plankton Research*, *26*(6), 625–635. https://doi.org/10.1093/plankt/fbh064

Spaak, P., Fox, J., & Hairston, N. G. (2012). Modes and mechanisms of a *Daphnia* invasion. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2936–2944. https://doi.org/10.1098/rspb.2012.0280

Spaak, P., & Hoekstra, J. R. (1995). Life History Variation and the Coexistence of a *Daphnia* Hybrid With Its Parental Species. *Ecology*, *76*(2), 553–564. https://doi.org/10.2307/1941213

Spaak, P., & Hoekstra, J. R. (1997). Fish predation on a *Daphnia* hybrid species complex: A factor explaining species coexistence? *Limnology and Oceanography*, *42*(4), 753–762. https://doi.org/10.4319/lo.1997.42.4.0753

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160455. https://doi.org/10.1098/rstb.2016.0455

Taylor, D. J., & Hebert, P. D. N. (1994). Genetic assessment of species boundaries in the North American *Daphnia longispina* complex (Crustacea: Daphniidae). *Zoological Journal of the Linnean Society*, *110*(1), 27–40. https://doi.org/10.1111/j.1096-3642.1994.tb01469.x

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177. https://doi.org/10.1038/s41559-018-0777-y

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

Tsugeki, N. K., Honjo, M. N., & Kuwae, M. (2021). Interspecific variation in ephippial size between *Daphnia galeata* and *D. pulicaria* in Lake Biwa, Japan. *Limnology*, *22*(2), 197–207. https://doi.org/10.1007/s10201-020-00646-8

Turner, L. M., & Harr, B. (2014). Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *eLife*, *3*, e02504. https://doi.org/10.7554/eLife.02504

Twyford, A. D., & Ennos, R. A. (2012). Next-generation hybridization and introgression. *Heredity*, *108*(3), 179–189. https://doi.org/10.1038/hdy.2011.68

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., Gordon, K. H. J., Specht, A., Paula-Moraes, S. V., Rane, R., Walsh, T. K., & Jiggins, C. D. (2020). Adaptive Introgression across Semipermeable Species Boundaries between Local *Helicoverpa zea* and Invasive *Helicoverpa armigera* Moths. *Molecular Biology and Evolution*, *37*(9), 2568–2583. https://doi.org/10.1093/molbev/msaa108

Vergilino, R., Markova, S., Ventura, M., Manca, M., & Dufresne, F. (2011). Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. *Molecular Ecology*, *20*(6), 1191–1207. https://doi.org/10.1111/j.1365-294X.2011.05004.x

Viard, F., Riginos, C., & Bierne, N. (2020). Anthropogenic hybridization at sea: Three evolutionary questions relevant to invasive species management. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *375*(1806), 20190547. https://doi.org/10.1098/rstb.2019.0547

Weider, L. J., & Stich, H. B. (1992). Spatial and temporal heterogeneity of *Daphnia* in Lake Constance; intra- and interspecific comparisons. *Limnology and Oceanography*, *37*(6), 1327–1334. https://doi.org/10.4319/lo.1992.37.6.1327

Wersebe, M. J., Sherman, R. E., Jeyasingh, P. D., & Weider, L. J. (2023). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology*, *32*(6), 1478–1496. https://doi.org/10.1111/mec.16383

Wersebe, M. J., & Weider, L. J. (2023). Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species. *Proceedings of the National Academy of Sciences*, *120*(6), e2217276120. https://doi.org/10.1073/pnas.2217276120

Westram, A. M., Faria, R., Johannesson, K., & Butlin, R. (2021). Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, *30*(15), 3797–3814. https://doi.org/10.1111/mec.15861

Whibley, A., Kelley, J. L., & Narum, S. R. (2021). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, *21*(3), 641–652. https://doi.org/10.1111/1755-0998.13312

Wolf, H. G. (1987). Interspecific hybridization between *Daphnia hyalina*, *D. galeata*, and *D. cucullata* and seasonal abundances of these species and their hybrids. *Hydrobiologia*, *145*(1), 213–217. https://doi.org/10.1007/BF02530282

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Wolinska, J., Bittner, K., Ebert, D., & Spaak, P. (2006). The coexistence of hybrid and parental *Daphnia*: The role of parasites. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1596), 1977–1983. https://doi.org/10.1098/rspb.2006.3523

Wolinska, J., Löffler, A., & Spaak, P. (2007). Taxon-specific reaction norms to predator cues in a hybrid *Daphnia* complex. *Freshwater Biology*, *52*(7), 1198–1209. https://doi.org/10.1111/j.1365-2427.2007.01757.x

Woodward, G., Perkins, D. M., & Brown, L. E. (2010). Climate change and freshwater ecosystems: Impacts across multiple levels of organization. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1549), 2093–2106. https://doi.org/10.1098/rstb.2010.0055

Wu, Z., Wang, Z., Xie, D., Zhang, J., Cai, P., Li, X., Xu, X., Li, T., & Zhao, J. (2022). Extensive Sympatry and Frequent Hybridization of Ecologically Divergent Aquatic Plants on the Qinghai-Tibetan Plateau. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.851151

Yang, W., Feiner, N., Laakkonen, H., Sacchi, R., Zuffi, M. A. L., Scali, S., While, G. M., & Uller, T. (2020). Spatial variation in gene flow across a hybrid zone reveals causes of reproductive isolation and asymmetric introgression in wall lizards. *Evolution*, *74*(7), 1289–1300. https://doi.org/10.1111/evo.14001

Ye, Z., Xu, S., Spitze, K., Asselman, J., Jiang, X., Ackerman, M. S., Lopez, J., Harker, B., Raborn, R. T., Thomas, W. K., Ramsdell, J., Pfrender, M. E., & Lynch, M. (2017). A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes, Genomes, Genetics*, *7*(5), 1405–1416. https://doi.org/10.1534/g3.116.038638

Yin, M., Gießler, S., Griebel, J., & Wolinska, J. (2014). Hybridizing *Daphnia* communities from ten neighbouring lakes: Spatio-temporal dynamics, local processes, gene flow and invasiveness. *BMC Evolutionary Biology*, *14*(1), 80. https://doi.org/10.1186/1471-2148-14-80

Yin, M., Wolinska, J., & Gießler, S. (2010). Clonal diversity, clonal persistence and rapid taxon replacement in natural populations of species and hybrids of the *Daphnia longispina* complex. *Molecular Ecology*, *19*(19), 4168–4178. https://doi.org/10.1111/j.1365-294X.2010.04807.x

Yousey, A. M., Chowdhury, P. R., Biddinger, N., Shaw, J. H., Jeyasingh, P. D., & Weider, L. J. (2018). Resurrected 'ancient' *Daphnia* genotypes show reduced thermal stress tolerance compared to modern descendants. *Royal Society Open Science*, *5*(3), 172193. https://doi.org/10.1098/rsos.172193

Zuykova, E. I., Bochkarev, N. A., Taylor, D. J., & Kotov, A. A. (2019). Unexpected endemism in the *Daphnia longispina* complex (Crustacea: Cladocera) in Southern Siberia. *PLoS One*, *14*(9), e0221527. https://doi.org/10.1371/journal.pone.0221527

Chapter 2 – Cost-saving population genomic investigation of *Daphnia longispina* complex resting eggs using whole-genome amplification and pre-sequencing screening

Jana Nickel[1], Mathilde Cordellier[1]

[1] Institute of Animal Cell and Systems Biology, University of Hamburg, Germany

This chapter is based on the published paper:

Nickel, J., & Cordellier, M. (2022). Cost-saving population genomic investigation of *Daphnia longispina* complex resting eggs using whole-genome amplification and pre-sequencing screening. *Ecology and Evolution*, *12*(12), e9682. https://doi.org/10.1002/ece3.9682

Abstract

Resting stages of aquatic organisms that accumulate in the sediment over time are an exceptional resource that allows direct insights into past populations and addressing evolutionary questions. This is of particular interest in taxa that face relatively new environmental challenges, e.g., climate change and eutrophication, such as the *Daphnia longispina* species complex, a keystone zooplankton group in European freshwater ecosystems. However, genomic analysis might be challenging as DNA yield from many of these resting stages can be low and the material degraded. To reliably allow the resequencing of single *Daphnia* resting eggs from different sediment layers and characterize genomic changes through time, we performed whole-genome amplification to obtain DNA amounts suitable for genome resequencing and tested multiple protocols involving egg isolation, whole-genome amplification kits, and library preparation. A pre-sequencing contamination screening was developed, consisting of amplifying mitochondrial *Daphnia* and bacterial markers, to quickly assess and exclude possibly contaminated samples. In total, we successfully amplified and sequenced nine genomes from *Daphnia* resting eggs that could be identified as *Daphnia longispina* species. We analyzed the genome coverage and heterozygosity of these samples to optimize this method for future projects involving population genomic investigation of the resting egg bank.

Keywords: *Daphnia longispina* species complex, population genomics, resting egg bank, whole-genome amplification

Introduction

Organisms that produce dormant propagules, e.g., seeds, eggs, cysts, or spores are especially interesting to investigate the evolutionary history of species, populations, and whole ecosystems. These dormant propagules accumulate in layers of limnic or marine sediment over time and function as a biological archive that allows direct insights into shifts in genetic variation and how past and ongoing environmental changes have shaped ecosystems (Hairston, 1996; Orsini et al., 2016). Combined with the rapidly growing genomic resources and high-throughput sequencing technologies, resting stage banks enabled researchers to investigate local adaptation in rotifers (Franch-Gras et al., 2018), genetic structure in diatoms (Härnström et al., 2011), and genetic diversity over time in *Daphnia magna* (Orsini et al., 2016). In freshwater sediment, resting egg banks are often dominated by zooplanktonic crustaceans such as the genus *Daphnia* (Crustacea, Cladocera) that play a key role in aquatic food webs; they graze on phytoplankton and are a food source for secondary consumers (Lampert & Sommer, 2007). As cyclical parthenogens, *Daphnia* are able to switch between asexual and sexual reproduction and the resulting resting eggs can withstand adverse conditions for decades and even centuries (Frisch et al., 2014). In some cases, resting eggs extracted from sediment cores can be hatched and clonal lines brought back to life to investigate temporal and spatial patterns in the recent past (Orsini et al., 2013). The DNA preserved in those resting eggs can also be directly analyzed with various molecular methods (Cousyn et al., 2001; Lack et al., 2018; Limburg & Weider, 2002) to study adaptation to changing environmental conditions such as temperature (Dziuba et al., 2020) or eutrophication (Alric et al., 2016; Cordellier et al., 2021).

Genomic investigation of the resting egg bank that can be conducted directly without hatching and establishing clonal lines is hindered by small amounts of potentially degraded DNA in single *Daphnia* resting eggs. Before high-throughput sequencing technologies became widespread in non-model organisms, population genetic studies were restricted to a few nuclear and mitochondrial markers to characterize resting eggs (Brede et al., 2009; Möst et al., 2015; Ortells et al., 2014). However, the small amount of tissue (~3500 cells per diapausing embryo in *D. magna*; Chen et al., 2018) makes whole-genome sequencing extremely challenging. One possible approach is pooling multiple eggs from a population for whole-genome sequencing but then information on individual genotypes is lost (Cordellier et al., 2021).

Another approach to obtain sufficient DNA from starting material that is of limited quantity and/or quality is multiple displacement amplification (MDA), a method for whole-genome amplification (WGA) commonly used to perform isothermal amplification of the template DNA.

MDA-WGA uses phi29 DNA polymerase and annealing of random hexamers which does not require species-specific primers and yields an average DNA product length of >10 kb (Dean et al., 2002; Spits et al., 2006). It is the preferred method for SNP detection (De Bourcy et al., 2014) and has been used in other studies where extremely small specimens hinder genomic investigation to successfully perform RADSeq (Cruaud et al., 2018) and whole-genome sequencing (O'Grady et al., 2022). These new methods enabled researchers to study introgression in Schistosome parasites (Platt et al., 2019) and population genomic structure in water mites (Blattner et al., 2022) and ghost-worms (Cerca et al., 2021). It can also be used to detect copy number variants (Deleye et al., 2017) and most structural variants (Lack et al., 2018). Potential drawbacks are increased cost for WGA kits and GC-dependent amplification bias (Sabina & Leamon, 2015).

In a previous study, Lack et al. (2018) demonstrated that it is possible to use WGA of *Daphnia pulicaria* resting eggs to achieve DNA concentration suitable for whole-genome sequencing but caution that it should only be used when necessary, e.g., when it is not possible to hatch eggs and sequence genomes of multiple clonal individuals. However, the hatching success of resting eggs is highly species-dependent, with *Daphnia magna* exhibiting a generally high hatching rate, and members of the *Daphnia longispina* species complex a poor hatching success. Further, within species, other factors such as lake of origin and sediment composition seem to play a role in hatching success (personal observation MC, Radzikowski et al., 2018). In addition, in populations of *D. longispina* species, hybrid resting eggs are less likely to hatch than their parental species in lab experiments (Schwenk et al., 2001) and natural populations (Keller et al., 2007; Keller & Spaak, 2004). Indeed, hybridization and introgression are common in the *D. longispina* species complex. This could result in bias towards the parental species when hatching and rearing clonal lines from resting eggs.

In this study, we tested multiple protocols involving egg isolation, WGA kits, and library preparation, and developed a contamination screening to reliably sequence genomes from single resting eggs from the *D. longispina* species complex. We also analyzed the read depth and heterozygosity to optimize this method for future projects involving population genomic investigation of the resting egg bank using recent and historical resting eggs.

Methods

Sampling and isolation

Sandy soil was collected by hand from the shoreline of the eutrophic lake Eichbaumsee, Germany (53° 29′ 6″ N, 10° 6′ 11″ E) and stored at 4°C. The exact age of the soil is unknown but the upper layers most likely contain recent *Daphnia* eggs from the last few years.

To collect *Daphnia* eggs small amounts of sediment were sieved (125 µm mesh size) and resuspended in ddH_2O. Ephippia were eye spotted, counted, and transferred to 1.5 ml tubes under a stereo microscope (Nikon SMZ800N). The water was then removed, and the samples were kept at 4°C in the dark until further processing. The ephippia were transferred to a drop of sterile 1× PBS and opened under a stereo microscope with insect needles and forceps previously treated with UV light in a PCR workstation Pro (VWR) and cleaned with DNA-ExitusPlus (PanReac AppliChem). If an egg was present a picture was taken and the quality was evaluated by eye based on their color into the categories light green or dark green, which is the highest quality we find. Eggs that had an already damaged egg membrane, an uneven shape or were orange were discarded (Marková et al., 2006). The resting egg separated from the ephippial casing was then transferred to a tube with sterile 1× PBS with a pipette to wash away the remaining material and the egg was transferred in 1 µl PBS to a new tube with 2, 3, or 14 µl fresh PBS, depending on the WGA protocol (REPLI-g Mini, Single cell and Single Cell increased sample volume, respectively). The isolated eggs were kept at −20 or −80°C at least overnight until amplification.

DNA extraction from batch cultures

As an unamplified control for the WGA samples, high-molecular-weight genomic DNA was extracted from 20 pooled adult *Daphnia* individuals (M5 clone; Nickel et al., 2021) using a modified CTAB extraction method as described in Cristescu et al. (2006).

Whole-genome amplification of isolated resting eggs

For whole-genome amplification of single eggs, the REPLI-g Single Cell Kit and REPLI-g Mini Kit (Qiagen) were used. Both kits are used for unbiased amplification of genomic loci due to MDA. The REPLI-g Single Cell Kit can be used for samples of 1–1000 intact cells and yields more DNA. The isolated resting eggs were thawed on ice and WGA was performed following the manufacturer's protocols. Briefly, denaturation buffer was added to the prepared resting eggs in PBS and amplified by phi29 DNA polymerase under isothermal conditions at 30°C for 8 h using the REPLI-g Single Cell Kit and 16 h using the REPLI-g Mini Kit and the polymerase was inactivated at 65°C for 3 min. In addition, a modified protocol for the REPLI-g Single Cell Kit as described by Lack et al. (2018) was used; it is optimized for the amplification of 10–100 ng genomic DNA template and uses an increased sample volume (15 µl).

Eggs were either kept intact or punctured with an insect needle before the amplification to test whether manual crushing had an effect. The different methods involving the REPLI-g kit, if the normal or increased sample volume protocol were used, storage, and egg integrity were performed on two eggs each for the nine tested protocol combinations and are shown in Table 2.1 and Figure 2.1.

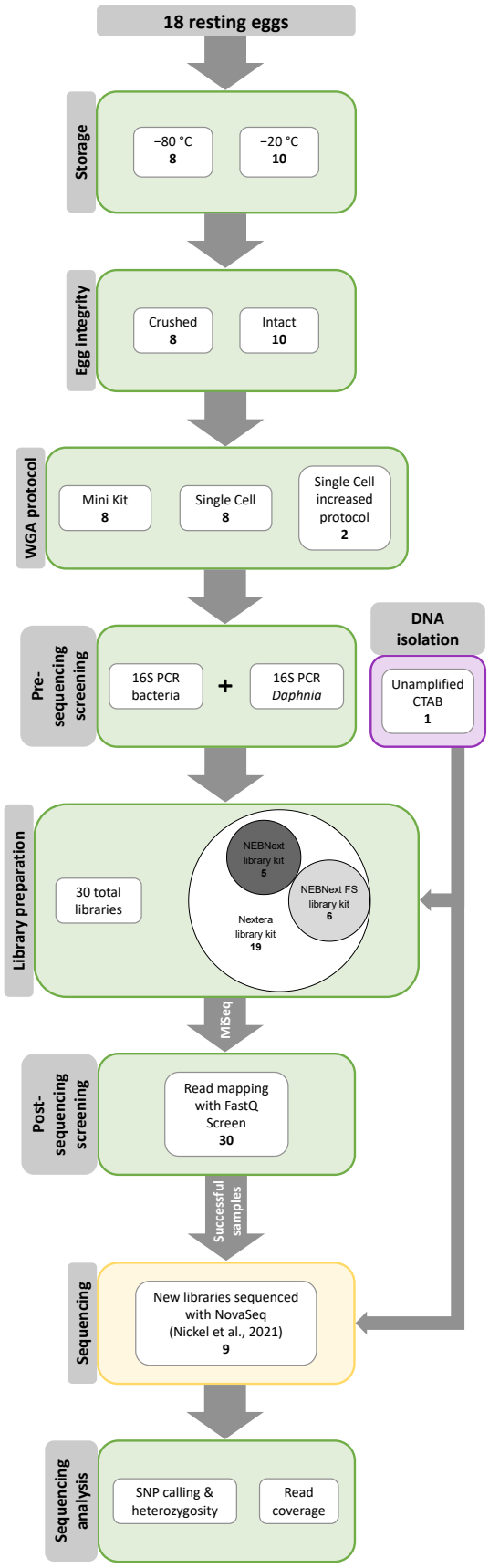


Figure 2.1. Experimental design workflow for multiple protocols of whole-genome amplification of isolated resting eggs, library preparation, sequencing, and bioinformatic analysis.

Table 2.1. Summary of all samples for egg quality, the complete method used for each egg, and PCR screening results (16S *Daphnia*/16S bacteria: Amplified product visible on a gel). For the sequencing statistics, the number of reads is presented separately for Illumina Nextera or NEB library preparations. The mapped reads % column shows the % of reads mapped to the *D. galeata* genome by BWA. Samples with very high mapping success are considered as successful amplification of *Daphnia* DNA and are indicated in bold while the remaining samples are considered contaminated.

Sample name	Egg quality	REPLI-g kit	Protocol	Storage	Egg integrity	16S *Daphnia*	16S bacteria	Total raw reads	Total trimmed reads	Mapped reads %
EIC_18	Dark	Single Cell	Normal	−20°C	intact	Yes	No	Nextera: 106,686	72,564	**93.64**
EIC_15.4	Light	Single Cell	Normal	−20°C	intact	Yes	Yes	Nextera: 55,146	36,468	11.53
EIC_2.1	Unknown	Single Cell	Increased sample	−20°C	intact	Yes	Weak	Nextera: 84,848	54,265	17.91
EIC_2.2	Unknown	Single Cell	Increased sample	−20°C	intact	Yes	Weak	Nextera: 64,268	41,923	20.58
EIC_12.3	Light	Single Cell	Normal	−80°C	intact	Weak	Weak	Nextera: 116,572	78,875	0.04
								NEB: 20,133	19,364	3.64
EIC_13.3	Dark	Single Cell	Normal	−80°C	intact	Yes	No	Nextera: 53,342	35,721	12.12
								NEB: 30,512	29,182	4.29
EIC_3.1	Light	Single Cell	Normal	−20°C	crushed	Yes	Weak	Nextera: 55,176	37,877	8.16
EIC_3.2	Dark	Single Cell	Normal	−20°C	crushed	No	Weak	Nextera: 43,025	26,622	13.82
EIC_16	Light	Single Cell	Normal	−80°C	crushed	Yes	No	Nextera: 156,382	99,335	**95.95**
								NEB FS: 62,699	49,802	**96.12**
EIC_17	Dark	Single Cell	Normal	−80°C	crushed	Yes	No	Nextera: 27,744	17,700	**96.69**
								NEB FS: 63,985	55,836	**96.25**
EIC_7.1	Dark	Mini	Normal	−20°C	intact	Yes	Weak	Nextera: 81,411	53,037	29.99
EIC_7.2	Dark	Mini	Normal	−20°C	intact	No	Yes	Nextera: 42,070	26,465	11.37
								NEB: 40,303	38,387	0.02
EIC_11	Dark	Mini	Normal	−80°C	intact	Yes	Weak	Nextera: 18,787	11,842	**98.06**
								NEB: 43,710	41,774	**96.74**
EIC_12	Light	Mini	Normal	−80°C	intact	Yes	No	Nextera: 63,732	38,445	**97.67**
								NEB: 40,089	38,354	**96.64**
EIC_13	Light	Mini	Normal	−20°C	crushed	Yes	No	Nextera: 102,521	64,463	**97.92**
EIC_13.2	Dark	Mini	Normal	−20°C	crushed	Yes	No	Nextera: 33	9	/
								NEB FS: 63,285	56,212	**97.99**
EIC_14	Light	Mini	Normal	−80°C	crushed	Yes	No	Nextera: 151,989	97,438	**98.23**
								NEB FS: 56,431	49,438	**97.64**
EIC_15	Light, slight dissolving	Mini	Normal	−80°C	crushed	Yes	Weak	Nextera: 89,600	57,955	**97.73**
								NEB FS: 47,563	41,308	**93.89**
M5	/	/	Unamplified CTAB	/	/	Yes	Weak	Nextera: 62,905	37,408	**92.45**
								NEB FS: 104,196	53,906	**93.89**

The amplified product was quantified on a NanoDrop spectrophotometer (ThermoFisher) to check that the A260/280 and A260/230 values were both >1.8 which indicates DNA purity. The concentration was measured with a Qubit Fluorometer (ThermoFisher) because during the REPLI-g reaction single-stranded DNA is generated by random extension of primer dimers which leads to an overestimation of DNA using a spectrophotometer. Successful amplification product was purified with 0.4× Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove small fragments and eluted in 60 µl 1× TE buffer. The cleaned genomic DNA was then quantified with a Qubit Fluorometer and fragment length was examined on a 4200 TapeStation (Agilent) or Fragment Analyzer (Agilent). The amplification product was stored at −20°C until library preparation. The presence of *Daphnia* DNA in the WGA product was checked by amplifying fragments of the mitochondrial 16S rDNA gene using the universal cladoceran primers S1 (5′-CGG CCG CCT GTT TAT CAA AAA CAT-3′) and S2 (5′-GGA GCT CCG GTT TGA ACT CAG ATC-3′) with 1 cycle of 93°C for 2 min 30 s, 55°C for 1 min and 72°C for 2 min followed by 40 cycles of 93°C for 1 min, 55°C for 1 min and 72°C for 2 min and running a 1.5% agarose gel at 100 V to assess bands (Schwenk et al., 1998). To check for a low presence of bacterial DNA universal primers for the bacterial 16S rDNA gene were used (5′-TCC TAC GGG AGG CAG CAG T-3′ and 5′-GGA CTA CCA GGG TAT CTA ATC CTG TT-3′) with 1 cycle of 50°C for 2 min and 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min (Nadkarni et al., 2002).

Library preparation and sequencing

Paired-end library construction was conducted for the 18 WGA samples and one unamplified CTAB sample with the Nextera XT DNA Library Preparation Kit (Illumina). Two library preparation kits were used on the same WGA samples to test which library kit could be best adapted to these samples. Five WGA samples (500 ng as input DNA) were fragmented using the M220 Focused-ultrasonicator (Covaris) and prepared with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs). In addition, five WGA samples (500 ng as input DNA) and one unamplified CTAB sample were prepared with NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina which includes an enzyme DNA fragmentation step. The obtained fragment length was measured prior to sequencing on a 4200 TapeStation (Agilent) with the High Sensitivity D5000 kit. The libraries were separately pooled and 150 bp paired-end reads were sequenced for the 19 Nextera libraries and 11 NEB libraries (30 total) on the Illumina MiSeq platform using the MiSeq Reagent Kit v2 Nano (Illumina).

Using only the eight WGA samples that were identified as largely *Daphnia* sequences in the previous low-coverage MiSeq sequencing step as well as one unamplified CTAB sample (9 total), new libraries were prepared with the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina. Then, 150 bp paired-end sequencing was generated on the Illumina NovaSeq 6000

platform as part of a previous study (Nickel et al., 2021). This whole-genome data was used here to have sufficient coverage to assess genome coverage and is available from the European Nucleotide Archive (accession numbers: ERR4610186–ERR4610192, ERR4610229, and ERR5235052). The successful sample EIC_13.2 could not be sequenced again because no amplification product was left.

Sequencing analysis and genotyping

The quality of raw and trimmed reads was assessed using FastQC v0.11.7 (Andrews, 2010). Trimming and quality filtering of the 30 total MiSeq datasets was performed using Trimmomatic v0.38 (Bolger et al., 2014) with the following parameters: TRAILING: 15 SLIDINGWINDOW: 4:15 MINLEN: 120. To assess contamination in the WGA samples FastQ Screen v0.14.0 with the BWA mapping option was used (Wingett & Andrews, 2018). A custom database was built to map trimmed reads against possible contaminants that included general common contaminants such as *Homo sapiens* (GRCh38.p7), the UniVec database, a bacterial and a viral NCBI reference set all downloaded in April 2018 as well as the *D. galeata* genome (Nickel et al., 2021), *D. magna* genome and *Acutodesmus obliquus* draft genome (Starkenburg et al., 2017) (Table S2.1). All sequences that did not map to any reference genome were filtered and searched against the GenBank nucleotide database (downloaded Feb. 2022) using blastn 2.12.0+ to identify their origin (Camacho et al., 2009). The trimmed reads were mapped to the *D. galeata* reference genome, using BWA v0.7.17 with the mem option (Li & Durbin, 2009) and Qualimap v2.2.1 was used to examine the mapping quality (Okonechnikov et al., 2016).

In addition, adapter trimming and quality filtering of the nine NovaSeq datasets from Nickel et al. (2021) were performed using Trimmomatic v0.38 with the following parameters: ILLUMINACLIP: TruSeq3-PE.fa: 2:30:10 TRAILING: 15 SLIDINGWINDOW: 4:15 MINLEN: 70. Following GATK4 best practices for pre-processing and variant calling (Van der Auwera et al., 2013) the trimmed reads were mapped to the *D. galeata* genome using BWA v0.7.17 with the mem and –M options (Li & Durbin, 2009). Duplicates were marked and filtered out in the BAM file using Picard v2.21.1 (http://broadinstitute.github.io/picard/).

We estimated read depth across the genome with a bin size of 10 kb and normalized with the Reads Per Kilobase per Million mapped reads (RPKM) model using bamCoverage from the deepTools package v3.5.1 (Ramírez et al., 2016).

To call variants for each sample HaplotypeCaller implemented in GATK v4.2.2.0 was run with the —emitRefConfidence GVCF and —include-non-variant-sites option (Poplin et al., 2018). This outputs gVCF files with information on all variant as well as invariant genotyped sites to be able to calculate the total number of genotyped sites within a genomic window. The VCF file was hard filtered to remove variants with a QualByDepth <10, StrandOddsRatio >3,

FisherStrand >60, mapping quality <40, MappingQualityRankSumTest <−8, and ReadPosRankSumTest <−5 and indels and multi-allelic sites were removed with BCFtools v1.9 (Li, 2011). The proportion of heterozygous sites of all genotyped sites was calculated using the Python script popgenWindows.py (github.com/simonhmartin/genomics_general release 0.3) with a sliding 250 kb window and a step size of 25 kb.

Results

Whole-genome amplification

The amplification step of genomes derived from resting eggs yielded 3.7–10.4 µg and 9.4–41 µg DNA per reaction for the REPLI-g Mini and REPLI-g Single Cell Kit, respectively, and generated very long fragments (fragment length peak 10,000–45,000 bp). Nineteen libraries prepared with Illumina Nextera and eleven libraries prepared with NEB were sequenced on the Illumina MiSeq platform, producing a total of 1,376,237 and 572,906 raw reads, respectively, with an average of 72,434 and 52,082 reads generated per library (Table 2.1). On average, 62.4% and 87.3% of the reads were retained after trimming and quality control.

Contamination and read mapping of MiSeq datasets

The trimmed reads were analyzed with FastQ Screen to assess possible contamination. We expected that for samples with successful whole-genome amplification, the majority of reads would map to *D. galeata* with little contamination from other genomes and show similar patterns to the unamplified sample M5. Ten samples indicated the presence of *Daphnia* DNA and mapped >75% to *D. galeata*, <1% to other organisms, and the remaining reads were unmapped or mapped to multiple genomes (Figure 2.2). However, five samples showed little to no reads mapped to the *D. galeata* genome (0–1%), contamination (~5%) with bacteria, and the remaining reads could not be mapped to a genome included in the panel (Figure 2.2). For three samples low amounts of reads mapped to the *D. galeata* genome (4–21%) and bacteria (~5%) and the sample EIC_13.3 showed significant human, bacterial and viral contamination most likely caused by contamination during lab work while handling the egg or whole-genome amplification. Additional BLAST searches of the large percentage of the reads that could not be mapped to any included reference genome in the potentially contaminated samples also revealed no substantial matches to other organisms.

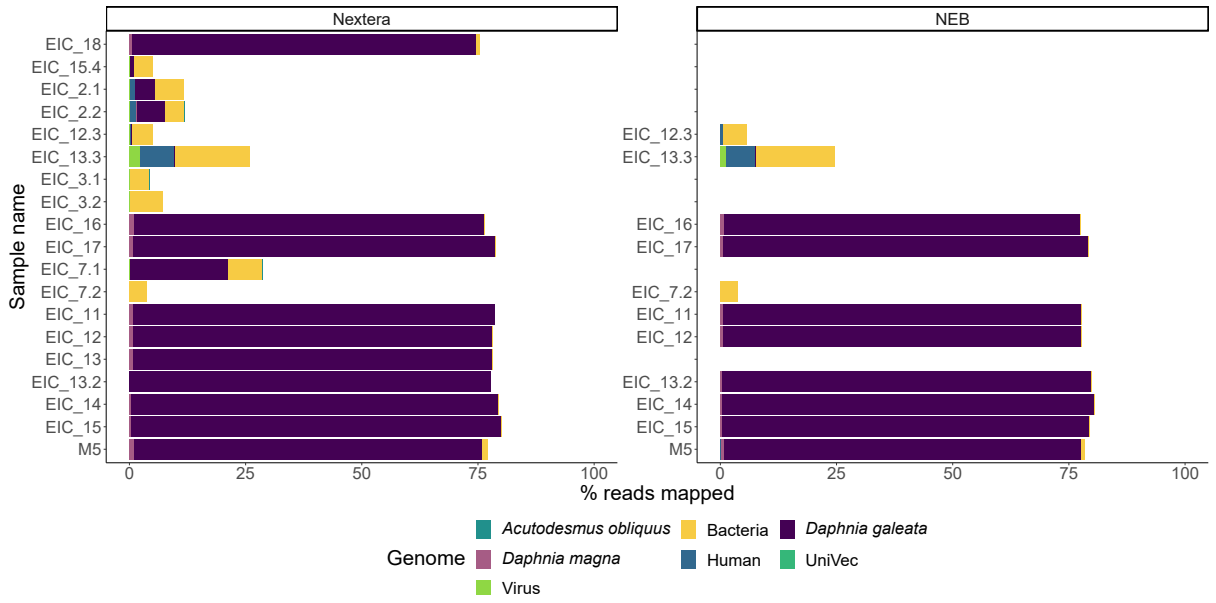


Figure 2.2. Proportion of reads mapped to a reference genome panel with FastQ Screen for all Illumina Nextera and NEB libraries. Reads that mapped to multiple genomes are not shown.

To properly analyze reads that mapped to multiple genomes with FastQ Screen and verify these results, the sequences were mapped separately to the *D. galeata* genome. We clearly identified *D. longispina* species with 92.5%–98.2% of reads from the successful ten samples mapping to the genome and average genome coverage of 0.1× (Table 2.1). For samples that were already identified by FastQ Screen as having no or small amounts of *D. galeata* reads only 0–30% of reads could be mapped. Samples, where both Nextera and NEB libraries were prepared, gave similar mapping rates and results are presented for both protocols in Figure 2.2. In total, nine *Daphnia* genomes from resting eggs were successfully amplified and sequenced which results in a 50% success rate as well as one *Daphnia* genome from a pooled unamplified DNA sample.

Effect of different protocols used

Out of the nine failed samples, only two were amplified using the REPLI-g Mini Kit while seven were amplified using the REPLI-g Single Cell Kit (normal and increased sample protocol). The latter yielded more DNA but produced lower quality WGA product and we were only able to successfully amplify samples using two protocol combinations (−20°C storage and leaving the eggs intact or −80°C storage and crushing the eggs). While separating resting eggs from the ephippial casing, only high-quality eggs that showed no sign of degeneration were selected for amplification and classified based on color but there does not seem to be a direct link between higher-quality dark green eggs and more amplification success. Out of the nine failed samples using either the REPLI-g Mini or Single Cell Kit, seven were stored at −20°C instead of −80°C.

Three eggs could be successfully amplified after keeping them intact before amplification and six that were manually crushed with an insect needle. However, because of their fragility, the

membrane of some of the intact eggs was most likely also punctured during isolation and transfer.

Three eggs out of ten tested eggs were successfully amplified using the REPLI-g Single Cell Kit, one using −20°C storage and leaving the eggs intact and two using −80°C storage and crushing the eggs. Six out of eight tested eggs could be amplified using the REPLI-g Mini Kit, two each using −80°C storage and either leaving the eggs intact or crushing them and two using −20°C storage and leaving the eggs intact.

The different library kits used generated a similar number of reads per library and a higher proportion of reads were retained after trimming using the NEB kits. In addition, the NEB kit yielded more consistent library concentrations and no failed libraries due to the protocol having more options to customize for different DNA input concentrations.

PCR contamination screening

Two different 16S PCR markers were used to assess the quality of the WGA product before sequencing and to compare these results to the results achieved by sequencing and mapping the reads. Sanger sequencing and BLAST search of the *Daphnia* 16S PCR fragments confirmed that all were of *D. galeata* or *D. longispina* mitochondrial origin. It is to be noted that Sanger sequencing was only conducted here for further diagnosis and is not necessary for the contamination screening.

Three samples had no or weak bands on the *Daphnia* 16S PCR gel and consequently failed during the sequencing. However, six samples where 16S was successfully amplified yielded no sequencing results (Table 2.1). All but one (EIC_13.3) of the samples that were classified as contaminated after sequencing showed a band during bacterial 16S PCR. However, this was also the case for two non-contaminated WGA samples and the unamplified control in which we would expect low amounts of bacterial DNA from the animals' microbiota. Combining both results and using only samples that pass *Daphnia* PCR and have no amplification of the bacterial DNA for subsequent sequencing could improve the success rate of WGA samples from 50% without any screening to 88% with stringent criteria. When considering more relaxed criteria that include all samples that passed the *Daphnia* PCR and with weak or no amplification of the bacterial DNA the success rate is 67%.

Read depth and coverage of NovaSeq datasets

While the MiSeq-generated data worked well to identify contamination quickly and at a lower cost, the low genome coverage (average 0.1×) was not sufficient to assess differences between unamplified and amplified samples and different WGA protocols used. To better compare the mapping rate between samples, assess the read coverage and variant calling,

we thus used the sequences of eight successful amplification samples and the unamplified, pooled sample with a higher coverage (0.60–57.05×, Table S2.2) obtained in another study (Nickel et al., 2021).

As MDA-WGA can lead to non-uniform amplification of the genome (Pinard et al., 2006) that could affect variant calling, we checked mapping success and distribution of read coverage across the genome. The mapping rate for all amplified samples was >90% which was similar to those calculated for the MiSeq datasets and the highest rates were achieved for the two samples using the REPLI-g Mini Kit, −80°C storage, and crushing the eggs (97.76% & 98.26%), the one sample using REPLI-g Mini Kit, −20°C storage and leaving the eggs intact (97.76%) and one of the samples using REPLI-g Mini Kit, −80°C storage and leaving the eggs intact (98.06%). The read coverage ranged from 5.55× to 12.41× for the amplified samples except for EIC_12 which showed extremely high adapter content. These were discarded during trimming and this subsequently resulted in lower coverage than the other samples (0.6×).

The normalized read depth for the unamplified sample M5 shows uniform coverage across the genome with few regions having low or very high coverage (Figure 2.3). In general, two samples that were prepared with the same protocol show very similar patterns of read coverage. The three samples that were prepared with the REPLI-g Single Cell Kit show a uniform coverage similar to the unamplified M5 sample, with most regions having a coverage above 5 and few regions with coverage above 50 that could point to overamplification of specific regions. For the samples that were amplified with the REPLI-g Mini Kit, most regions have coverage above 1 but some regions show much lower or higher coverage, especially sample EIC_14.

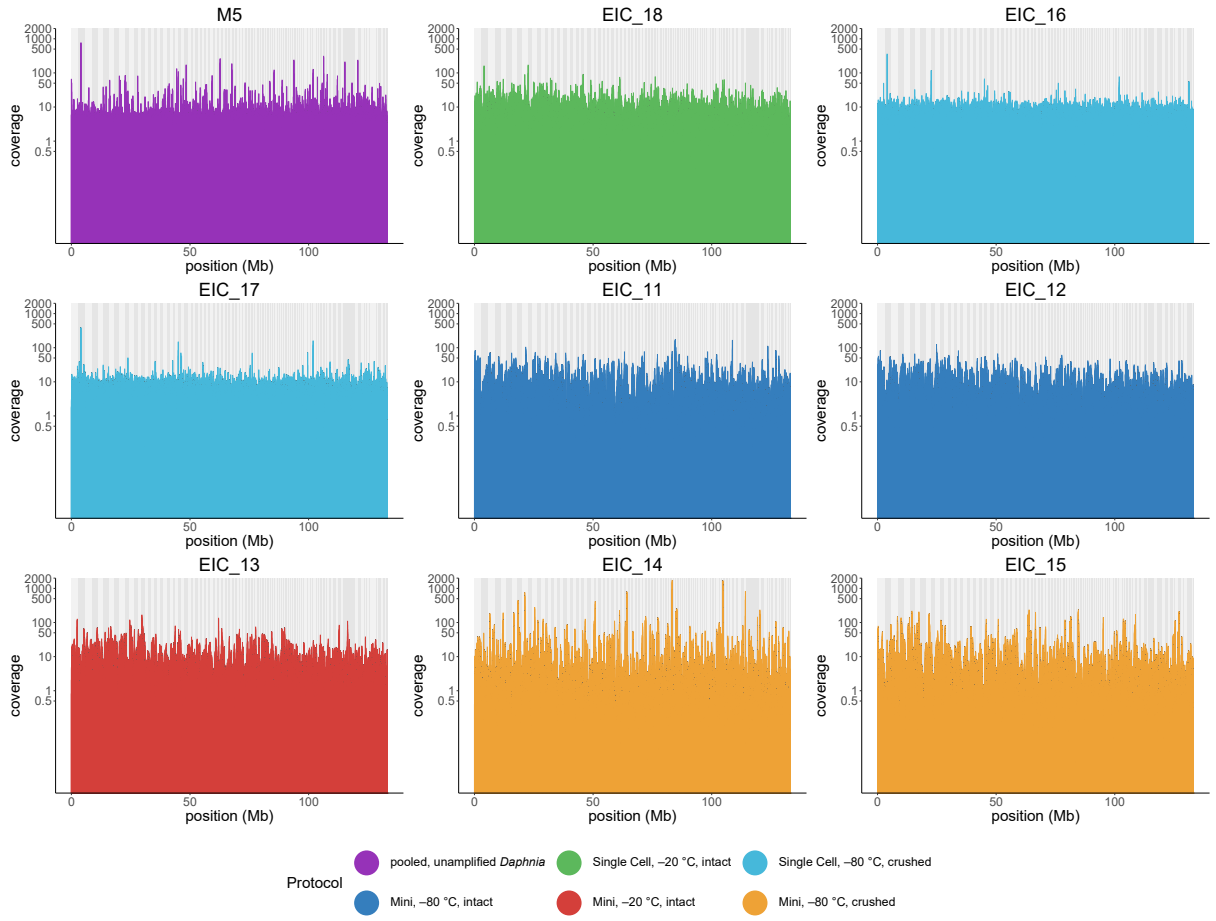


Figure 2.3. Normalized read depth across the genome in 10 kb bins for each sample. Scaffold positions are marked by alternating gray background bars. Coverage is shown on a log2 scale.

Variant calling and heterozygosity

The invariant data set included 130,950,194 sites across the nine samples. To assess whether we find downwardly biased estimates of heterozygosity in the amplified samples we compared the proportion of heterozygous genotype calls in sliding windows across the genome to the unamplified sample M5 (Figure 2.4). The heterozygosity across the genome in the unamplified sample was even with few outlier windows and the genome-wide average heterozygosity was 0.00934. We find very similar genome-wide patterns for the eight amplified samples and no general trend of loss of heterozygosity, with three samples having higher average heterozygosity compared with M5 and five having lower heterozygosity (Table S2.3). However, as the resting eggs were sampled from a natural population some differences in heterozygosity between two samples using the same protocol are expected.

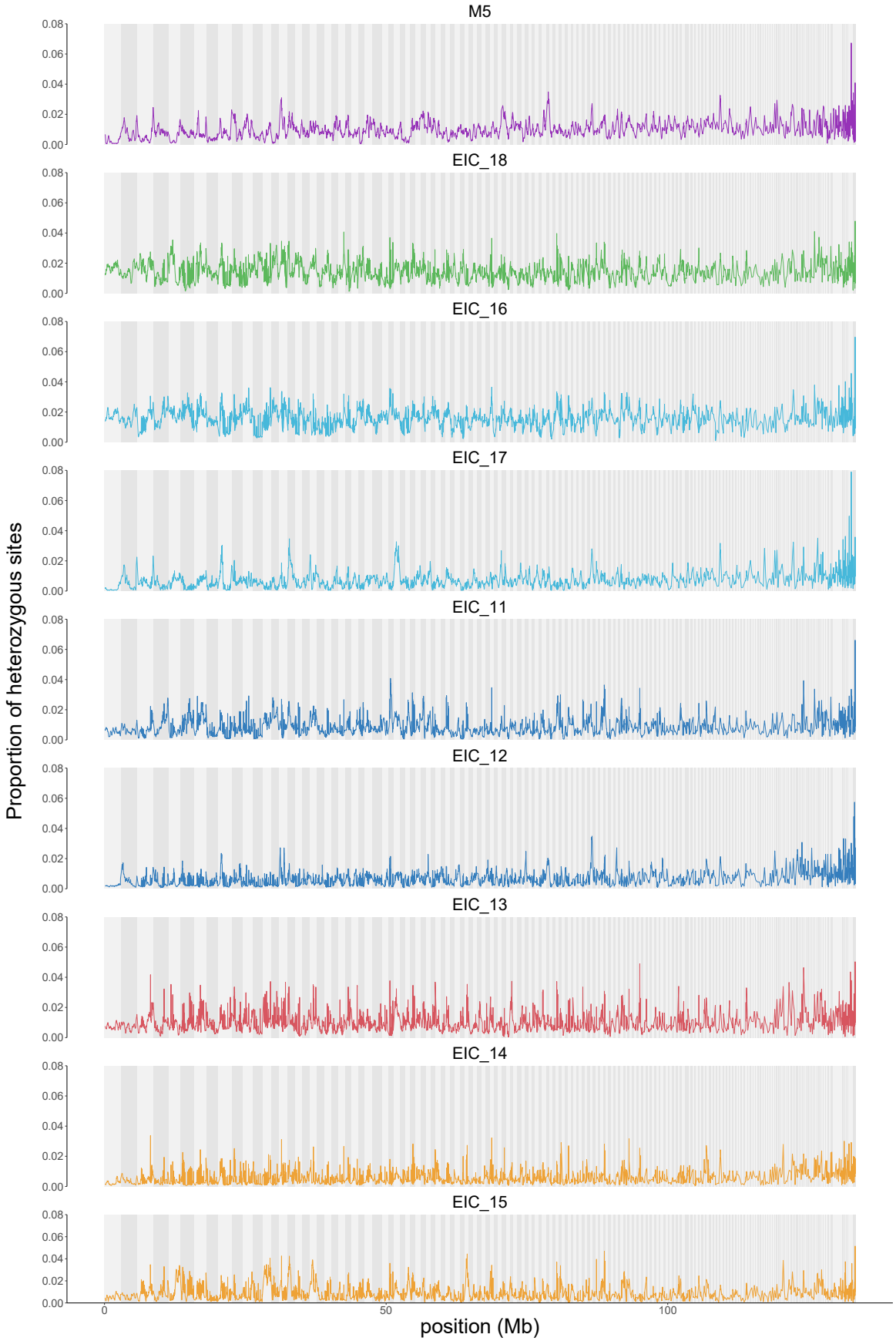


Figure 2.4. The proportion of heterozygous genotype calls in 250 kb sliding windows across the genome. Scaffold positions are marked by alternating gray background bars. Sample M5 is unamplified and the remaining eight are WGA samples using the same color code for the specific protocol as in Figure 2.3.

Discussion

Egg banks of zooplankton allow researchers to track long-term genetic and ecological variation within an ecosystem and provide insight into past populations by hatching long-dormant eggs or using genetic markers (Alric et al., 2016; Burge et al., 2018; Frisch et al., 2014). However, these studies are still limited by the reduction in egg viability with sediment age and the low amount of high-quality DNA in the eggs. The goal of this study was to optimize the whole-genome sequencing of *D. longispina* species resting eggs and establish a more reliable WGA method using 18 single resting eggs isolated from sediment. Hatching eggs from the resting egg bank and establishing clonal lines in the lab can be unpredictable and success rates depend on the species, the water bodies where sediment was collected, and the hatching conditions used (Radzikowski et al., 2018). In addition, *D. longispina* species hybrid resting eggs show lower hatching success and survival rates than their parental species (Schwenk et al., 2001). This introduces bias where hybrids appear less frequent than they are when working on admixed populations from the resting egg bank. We suggest that our method could help get more accurate genomic data from hybrid populations in other studies (Nickel et al., 2021).

In our study, we included the WGA method presented in Lack et al. (2018) using *Daphnia pulicaria* resting eggs where they were able to sequence one out of three resting eggs. We did not achieve successful amplification for the two tested *D. galeata* eggs using the same protocol (EIC_2.1 and EIC_2.2). In addition, Lack et al. (2018) did not use any qualitative diagnostic to assess the amplified *Daphnia* DNA from a resting egg before performing costly library preparation and sequencing steps. The success rate achieved by O'Grady et al. (2022) was much higher for eggs of *Daphnia magna* and *Daphnia pulicaria* (86% and 78%, respectively). Their study also did not include any screening step and so far, no species of the species predominant in European lakes were tested.

Here, different protocols for egg treatment and different library kits were used to successfully amplify and sequence nine genomes from *Daphnia* resting eggs that could be identified as belonging to the *D. longispina* species complex. *Daphnia* and bacterial markers were established to quickly check possible contamination of the WGA product prior to sequencing and exclude contaminated samples before the sequencing step at a low cost, thus improving the success rate to 88%. This may be of particular importance because the success rate can vary across different lakes and may be very low in specific lakes because of the poor condition of resting eggs (Marková et al., 2006). The contamination screening helps to identify potentially successful amplification of *Daphnia* DNA before the library preparation and sequencing and generate genomic data under these difficult conditions. The bacterial markers can be used to identify bacterial contamination in WGA of all other species, while the *Daphnia* markers work

for all Cladocera species, many of which also produce resting stages (Vandekerkhove et al., 2005). Our method could be improved with the use of real-time PCR to directly quantify *Daphnia* DNA in the sample instead of only examining whether it is detectable or not, albeit increasing the cost per sample.

To check for possible amplification bias, genome coverage and heterozygosity from amplified samples were compared with an unamplified DNA sample from pooled adult *Daphnia*. The read mapping ratio was high for all amplified samples and the genome coverage is relatively even with few outlier regions and very similar patterns for samples using the same protocol. Loss of heterozygosity that could impact genotyping was not observed in all samples. While the REPLI-g Single Cell Kit generally achieves good results for those metrics the low reproducibility of the amplification makes it unsuitable for our purpose.

Some challenges with whole-genome amplification were contamination which most likely stemmed from problems during the amplification step. Lab protocols to minimize contamination were used but the high sensitivity of the WGA kits could lead to the amplification of DNA from the wrong cells present in the sample.

A critical step is to use high-quality undamaged eggs that show no signs of degradation. In eggs from older sediment layers, this is often more difficult and will be tested in a different study. The benefits of this method included the potential to go back decades to centuries because DNA is preserved longer than eggs that can be reliably hatched (Limburg & Weider, 2002). However, some amplifications of resting eggs failed but we were not able to identify one or multiple specific organisms as major contaminants and most reads could not be mapped. Instead, we hypothesize that these unknown sequences could be caused by the phi29 polymerase performing non-templated DNA synthesis which is a known phenomenon in MDA-WGA and produces “junk” DNA possibly when the egg is already degrading or the amount of DNA present is too small (Nelson, 2014). This seems to be a more frequent problem using the REPLI-g Single Cell Kit where only three out of ten samples were successfully amplified.

In conclusion, the most appropriate complete protocol we tested included using the REPLI-g Mini Kit, storing eggs at −80°C, leaving the eggs intact, and using the NEB library kit. After the resting egg is removed from the ephippia, it is extremely fragile, and immediately freezing it at −80°C seems to be a crucial step to slow DNA degradation. Nevertheless, the high biological variability of the resting eggs and the relatively small number of eggs tested for each protocol makes it difficult to draw more general conclusions.

The possible shortcomings of WGA methods include the amplification of contaminant DNA instead or in addition to the template DNA (Thoendel et al., 2017), and increased costs for

sample preparation. Currently, the cost for the suggested REPLI-g Mini Kit is ~$8 per sample. The amplification of contaminant DNA remains an issue; however, our contamination screening when applied at a larger scale would lead to substantial cost savings, by markedly reducing the number of contaminated samples being processed further and sequenced. While sequencing itself has become extremely low-cost, library preparation remains costly. Prices for reagents, kits, labor, and sequencing services vary considerably between countries and are further influenced by the scale of purchasing. We, therefore, refrain from providing exact cost calculations. When using WGA strategies, it is also important to consider the impact of the quantity of input DNA that can lead to downwardly biased estimates of heterozygosity and therefore genotyping bias (De Medeiros & Farrell, 2018). This study and others that use very small amounts of input DNA (Campbell et al., 2020; Cruaud et al., 2018; O'Grady et al., 2022) indicate that MDA-WGA does not introduce amplification bias that affects SNP genotyping. It is also suitable for structural variant calling with the exception of inversions (Lack et al., 2018).

To sum it up, our method will allow the resequencing of resting eggs from different sediment layers to characterize genomic changes through time in the *D. longispina* species complex. In a broader context, WGA could also be used for resting stages of other organisms with low amounts of DNA such as other Cladocera taxa, rotifers, or diatoms to gain a more complete understanding of freshwater ecosystems.

Data accessibility

The NovaSeq short read data is available in the European Nucleotide Archive under run accession numbers ERR4610186–ERR4610192, ERR4610229, and ERR5235052, and the MiSeq datasets are deposited in Zenodo (https://doi.org/10.5281/zenodo.7253120).

Acknowledgments

We would like to thank Fynn Eilers and Jennifer Drechsler for their support in the molecular laboratory and with sampling. We are grateful to the Fischer group at the Institute of Food Chemistry (University of Hamburg) for providing us access to the Illumina MiSeq instrument. Open Access funding enabled and organized by Projekt DEAL.

Chapter 2 References

Alric, B., Möst, M., Domaizon, I., Pignol, C., Spaak, P., & Perga, M.-E. (2016). Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology*, *29*(4), 720–735. https://doi.org/10.1111/jeb.12820

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Blattner, L., Lucek, K., Beck, N., Berner, D., & Von Fumetti, S. (2022). Intra-Alpine Islands: Population genomic inference reveals high degree of isolation between freshwater spring habitats. *Diversity and Distributions*, *28*(2), 291–305. https://doi.org/10.1111/ddi.13461

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences*, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Burge, D. R. L., Edlund, M. B., & Frisch, D. (2018). Paleolimnology and resurrection ecology: The future of reconstructing the past. *Evolutionary Applications*, *11*(1), 42–59. https://doi.org/10.1111/eva.12556

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Campbell, E. O., Dupuis, J. R., Holowachuk, J., Hladun, S., Vankosky, M. A., & Mori, B. A. (2020). Disjunction between canola distribution and the genetic structure of its recently described pest, the canola flower midge (*Contarinia brassicola*). *Ecology and Evolution*, *10*(23), 13284–13296. https://doi.org/10.1002/ece3.6927

Cerca, J., Rivera-Colón, A. G., Ferreira, M. S., Ravinet, M., Nowak, M. D., Catchen, J. M., & Struck, T. H. (2021). Incomplete lineage sorting and ancient admixture, and speciation without morphological change in ghost-worm cryptic species. *PeerJ*, *9*, e10896. https://doi.org/10.7717/peerj.10896

Chen, L., Barnett, R. E., Horstmann, M., Bamberger, V., Heberle, L., Krebs, N., Colbourne, J. K., Gómez, R., & Weiss, L. C. (2018). Mitotic activity patterns and cytoskeletal changes throughout the progression of diapause developmental program in *Daphnia*. *BMC Cell Biology*, *19*(1), 30. https://doi.org/10.1186/s12860-018-0181-0

Cordellier, M., Wojewodzic, M. W., Wessels, M., Kuster, C., & Von Elert, E. (2021). Next-generation sequencing of DNA from resting eggs: Signatures of eutrophication in a lake's sediment. *Zoology*, *145*, 125895. https://doi.org/10.1016/j.zool.2021.125895

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D., & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, *98*(11), 6256–6260. https://doi.org/10.1073/pnas.111606798

Cristescu, M. E. A., Colbourne, J. K., Radivojac, J., & Lynch, M. (2006). A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex*: On the prospect of crustacean genomics. *Genomics*, *88*(4), 415–430. https://doi.org/10.1016/j.ygeno.2006.03.007

Cruaud, A., Groussier, G., Genson, G., Sauné, L., Polaszek, A., & Rasplus, J.-Y. (2018). Pushing the limits of whole genome amplification: Successful sequencing of RADseq library from a single microhymenopteran (Chalcidoidea, *Trichogramma*). *PeerJ*, *6*, e5640. https://doi.org/10.7717/peerj.5640

De Bourcy, C. F. A., Vlaminck, I. D., Kanbar, J. N., Wang, J., Gawad, C., & Quake, S. R. (2014). A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *PLoS One*, *9*(8), e105585. https://doi.org/10.1371/journal.pone.0105585

De Medeiros, B. A. S., & Farrell, B. D. (2018). Whole-genome amplification in double-digest RADseq results in adequate libraries but fewer sequenced loci. *PeerJ*, *6*, e5089. https://doi.org/10.7717/peerj.5089

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences*, *99*(8), 5261–5266. https://doi.org/10.1073/pnas.082089499

Deleye, L., Tilleman, L., Vander Plaetsen, A.-S., Cornelis, S., Deforce, D., & Van Nieuwerburgh, F. (2017). Performance of four modern whole genome amplification methods for copy number variant detection in single cells. *Scientific Reports*, *7*(1), 3422. https://doi.org/10.1038/s41598-017-03711-y

Dziuba, M. K., Herdegen-Radwan, M., Pluta, E., Wejnerowski, Ł., Szczuciński, W., & Cerbin, S. (2020). Temperature increase altered *Daphnia* community structure in artificially heated lakes: A potential scenario for a warmer future. *Scientific Reports*, *10*(1), 13956. https://doi.org/10.1038/s41598-020-70294-6

Franch-Gras, L., Hahn, C., García-Roger, E. M., Carmona, M. J., Serra, M., & Gómez, A. (2018). Genomic signatures of local adaptation to the degree of environmental predictability in rotifers. *Scientific Reports*, *8*(1), 16051. https://doi.org/10.1038/s41598-018-34188-y

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368. https://doi.org/10.1111/ele.12237

Hairston, N. G. (1996). Zooplankton egg banks as biotic reservoirs in changing environments. *Limnology and Oceanography*, *41*(5), 1087–1092.

Härnström, K., Ellegaard, M., Andersen, T. J., & Godhe, A. (2011). Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences*, *108*(10), 4252–4257. https://doi.org/10.1073/pnas.1013528108

Keller, B., & Spaak, P. (2004). Nonrandom sexual reproduction and diapausing egg production in a *Daphnia* hybrid species complex. *Limnology and Oceanography*, *49*(4, part 2), 1393–1400. https://doi.org/10.4319/lo.2004.49.4_part_2.1393

Keller, B., Wolinska, J., Tellenbach, C., & Spaak, P. (2007). Reproductive isolation keeps hybridizing *Daphnia* species distinct. *Limnology and Oceanography*, *52*(3), 984–991. https://doi.org/10.4319/lo.2007.52.3.0984

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2018). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, *18*(1), 118–127. https://doi.org/10.1111/1755-0998.12720

Lampert, W., & Sommer, U. (2007). *Limnoecology: The Ecology of Lakes and Streams* (Second Edition). Oxford University Press.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Limburg, P. A., & Weider, L. J. (2002). 'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1488), 281–287. https://doi.org/10.1098/rspb.2001.1868

Marková, S., Černý, M., Rees, D. J., & Stuchlík, E. (2006). Are they still viable? Physical conditions and abundance of *Daphnia pulicaria* resting eggs in sediment cores from lakes in the Tatra Mountains. *Biologia*, *61*(18), S135–S146. https://doi.org/10.2478/s11756-006-0126-5

Möst, M., Oexle, S., Marková, S., Aidukaite, D., Baumgartner, L., Stich, H.-B., Wessels, M., Martin-Creuzburg, D., & Spaak, P. (2015). Population genetic dynamics of an invasion reconstructed from the sediment egg bank. *Molecular Ecology*, *24*(16), 4074–4093. https://doi.org/10.1111/mec.13298

Nadkarni, M. A., Martin, F. E., Jacques, N. A., & Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, *148*(1), 257–266. https://doi.org/10.1099/00221287-148-1-257

Nelson, J. R. (2014). Random-Primed, Phi29 DNA Polymerase-Based Whole Genome Amplification. *Current Protocols in Molecular Biology*, *105*(1), 15.13.1–15.13.16. https://doi.org/10.1002/0471142727.mb1513s105

Nickel, J., Schell, T., Holtzem, T., Thielsch, A., Dennis, S. R., Schlick-Steiner, B. C., Steiner, F. M., Möst, M., Pfenninger, M., Schwenk, K., & Cordellier, M. (2021). Hybridization Dynamics and Extensive Introgression in the *Daphnia longispina* Species Complex: New Insights from a High-Quality *Daphnia galeata* Reference Genome. *Genome Biology and Evolution*, *13*(12), evab267. https://doi.org/10.1093/gbe/evab267

O'Grady, C. J., Dhandapani, V., Colbourne, J. K., & Frisch, D. (2022). Refining the evolutionary time machine: An assessment of whole genome amplification using single historical *Daphnia* eggs. *Molecular Ecology Resources*, *22*(3), 946–961. https://doi.org/10.1111/1755-0998.13524

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Orsini, L., Marshall, H., Cuenca Cambronero, M., Chaturvedi, A., Thomas, K. W., Pfrender, M. E., Spanier, K. I., & De Meester, L. (2016). Temporal genetic stability in natural populations of the waterflea *Daphnia magna* in response to strong selection pressure. *Molecular Ecology*, *25*(24), 6024–6038. https://doi.org/10.1111/mec.13907

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J. K., Pfrender, M. E., & Weider, L. J. (2013). The evolutionary time machine: Using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution*, *28*(5), 274–282. https://doi.org/10.1016/j.tree.2013.01.009

Ortells, R., Vanoverbeke, J., Louette, G., & De Meester, L. (2014). Colonization of *Daphnia magna* in a newly created pond: Founder effects and secondary immigrants. *Hydrobiologia*, *723*(1), 167–179. https://doi.org/10.1007/s10750-013-1593-7

Pinard, R., De Winter, A., Sarkis, G. J., Gerstein, M. B., Tartaro, K. R., Plant, R. N., Egholm, M., Rothberg, J. M., & Leamon, J. H. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*, *7*(1), 216. https://doi.org/10.1186/1471-2164-7-216

Platt, R. N., II, McDew-White, M., Le Clec'h, W., Chevalier, F. D., Allan, F., Emery, A. M., Garba, A., Hamidou, A. A., Ame, S. M., Webster, J. P., Rollinson, D., Webster, B. L., & Anderson, T. J. C. (2019). Ancient Hybridization and Adaptive Introgression of an Invadolysin Gene in Schistosome Parasites. *Molecular Biology and Evolution*, *36*(10), 2127–2142. https://doi.org/10.1093/molbev/msz154

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. https://doi.org/10.1101/201178

Radzikowski, J., Krupińska, K., & Ślusarczyk, M. (2018). Different thermal stimuli initiate hatching of *Daphnia* diapausing eggs originating from lakes and temporary waters. *Limnology*, *19*(1), 81–88. https://doi.org/10.1007/s10201-017-0520-4

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

Sabina, J., & Leamon, J. H. (2015). Bias in Whole Genome Amplification: Causes and Considerations. In T. Kroneis (Ed.), *Whole Genome Amplification: Methods and Protocols* (pp. 15–41). Springer. https://doi.org/10.1007/978-1-4939-2990-0_2

Schwenk, K., Bijl, M., & Menken, S. B. J. (2001). Experimental interspecific hybridization in *Daphnia*. *Hydrobiologia*, *442*(1), 67–73. https://doi.org/10.1023/A:1017594325506

Schwenk, K., Sand, A., Boersma, M., Brehm, M., Mader, E., Offerhaus, D., & Spaak, P. (1998). Genetic markers, genealogies and biogeographic patterns in the cladocera. *Aquatic Ecology*, *32*(1), 37–51. https://doi.org/10.1023/A:1009939901198

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., & Sermon, K. (2006). Whole-genome multiple displacement amplification from single cells. *Nature Protocols*, *1*(4), 1965–1970. https://doi.org/10.1038/nprot.2006.326

Starkenburg, S. R., Polle, J. E. W., Hovde, B., Daligault, H. E., Davenport, K. W., Huang, A., Neofotis, P., & McKie-Krisberg, Z. (2017). Draft Nuclear Genome, Complete Chloroplast Genome, and Complete Mitochondrial Genome for the Biofuel/Bioproduct Feedstock Species *Scenedesmus obliquus* Strain DOE0152z. *Genome Announcements*, *5*(32), e00617-17. https://doi.org/10.1128/genomeA.00617-17

Thoendel, M., Jeraldo, P., Greenwood-Quaintance, K. E., Yao, J., Chia, N., Hanssen, A. D., Abdel, M. P., & Patel, R. (2017). Impact of Contaminating DNA in Whole-Genome Amplification Kits Used for Metagenomic Shotgun Sequencing for Infection Diagnosis. *Journal of Clinical Microbiology*, *55*(6), 1789–1801. https://doi.org/10.1128/JCM.02402-16

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vandekerkhove, J., Declerck, S., Brendonck, L., Conde-Porcuna, J. M., Jeppesen, E., & De Meester, L. (2005). Hatching of cladoceran resting eggs: Temperature and photoperiod. *Freshwater Biology*, *50*(1), 96–104. https://doi.org/10.1111/j.1365-2427.2004.01312.x

Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, *7*, 1338. https://doi.org/10.12688/f1000research.15931.1

Chapter 3 – Hybridization dynamics and extensive introgression in the *Daphnia longispina* species complex: new insights from a high-quality *Daphnia galeata* reference genome

Jana Nickel[1], Tilman Schell[2], Tania Holtzem[3], Anne Thielsch[4], Stuart R. Dennis[5], Birgit C. Schlick-Steiner[3], Florian M. Steiner[3], Markus Möst[3], Markus Pfenninger[2,6,7], Klaus Schwenk[4], and Mathilde Cordellier[1]

[1] Institute of Zoology, Universität Hamburg, Germany

[2] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany

[3] Department of Ecology, University of Innsbruck, Austria

[4] Molecular Ecology, Institute for Environmental Sciences, University Koblenz-Landau, Landau in der Pfalz, Germany

[5] Department of Aquatic Ecology, EAWAG, Dübendorf, Switzerland

[6] Molecular Ecology, Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany

[7] IoME, Gutenberg University, Mainz, Germany

This chapter is based on the published paper:

Nickel, J., Schell, T., Holtzem, T., Thielsch, A., Dennis, S. R., Schlick-Steiner, B. C., Steiner, F. M., Möst, M., Pfenninger, M., Schwenk, K., & Cordellier, M. (2021). Hybridization Dynamics and Extensive Introgression in the *Daphnia longispina* Species Complex: New Insights from a High-Quality *Daphnia galeata* Reference Genome. *Genome Biology and Evolution*, *13*(12), evab267. https://doi.org/10.1093/gbe/evab267

Abstract

Hybridization and introgression are recognized as an important source of variation that influence adaptive processes; both phenomena are frequent in the genus *Daphnia*, a keystone zooplankton taxon in freshwater ecosystems that comprises several species complexes. To investigate genome-wide consequences of introgression between species, we provide here the first high-quality genome assembly for a member of the *Daphnia longispina* species complex, *Daphnia galeata*. We further resequenced 49 whole genomes of three species of the complex and their interspecific hybrids both from genotypes sampled in the water column and from single resting eggs extracted from sediment cores. Populations from habitats with diverse ecological conditions offered an opportunity to study the dynamics of hybridization linked to ecological changes and revealed a high prevalence of hybrids. Using phylogenetic and population genomic approaches, we provide first insights into the intra- and interspecific genome-wide variability in this species complex and identify regions of high divergence. Finally, we assess the length of ancestry tracts in hybrids to characterize introgression patterns across the genome. Our analyses uncover a complex history of hybridization and introgression reflecting multiple generations of hybridization and backcrossing in the *Daphnia longispina* species complex. Overall, this study and the new resources presented here pave the way for a better understanding of ancient and contemporary gene flow in the species complex and facilitate future studies on resting egg banks accumulating in lake sediment.

Keywords: introgression, hybridization, resting eggs, species complex, whole-genome amplification, genome assembly

Significance

Hybridization and introgression frequently occur in the water flea genus *Daphnia*. Until now, genomic resources to investigate the amount of introgression in the *Daphnia longispina* species complex, a widespread taxon in European lakes and keystone grazer of phytoplankton, were limited. We provide here a genome assembly and the first genome-wide analysis of several species of the complex, based on samples from the water column and resting eggs from sediment cores. Using phylogenetic and population genomic approaches, we measure intra- and interspecific genome-wide variability and identify regions of high divergence. Our study reveals extensive gene flow between taxa, and later-generation hybrids originating from several rounds of backcrossing. It paves the way for further studies aiming at understanding how species boundaries are maintained in the face of gene flow.

Introduction

Gene flow between species can be pervasive and can affect substantial parts of the genome. Hybridization and introgression are recognized as an important source of variation that can influence adaptive processes in plants, animals, yeast, and fungi (reviewed in Abbott et al., 2013; Arnold & Martin, 2009). The amount of realized gene flow varies among taxa and along the genome; it is governed by intrinsic genomic features such as recombination rate, structural variation, and intrinsic incompatibilities, as well as the species' biology and ecology including ecological and sexual selection, migration, and mode of reproduction.

How can species in diversifying clades frequently hybridize and show introgression but nevertheless maintain species boundaries? A growing body of literature provides examples for a high variety of systems where speciation occurs in the face of gene flow (Fraïsse et al., 2014; Martin et al., 2019; Meier et al., 2017). However, it is important to recognize that these systems are distributed along a wide spectrum. On one side of this spectrum, hybridization occurs but is not followed by introgression for several reasons such as reduced hybrid fertility or strong selection against hybrid phenotypes, leading to rapid hybrid breakdown. Barth et al. (2020) found that species boundaries in tropical eels are stable despite millions of years of hybridization, and also observed very few admixed individuals beyond F1 and first-generation backcrosses. The hybrid breakdown observed in this system reduces the likelihood of introgression via backcrossing. On the other side of the spectrum, hybridization is followed by introgression, and ongoing exchange of genetic information between species (Butlin et al., 2014; Doellman et al., 2018; Kaiser et al., 2021; Martin et al., 2013). Several empirical studies (Canestrelli et al., 2017; Schreiber & Pfenninger, 2021) as well as theoretical models (Flaxman et al., 2014; Rafajlović et al., 2016; Yeaman & Whitlock, 2011) suggest the possibility of intermediate constant equilibrium states, meaning that certain parts of the genome remain diverged (islands or continents of divergence), whereas others are freely exchanged among closely related species without ever reaching complete genomic isolation.

Recurrent hybridization and introgression are frequent in the genus *Daphnia* (Crustacea, Cladocera). Members of the genus have served as ecological model organisms for over a century (Miner et al., 2012), and the first crustacean genome to be sequenced was that of a member of the *Daphnia pulex* species complex (Colbourne et al., 2011). Since then, the genomes of 45 crustaceans have been sequenced with a focus on species of economic or medical interest (NCBI, last accessed January 2021). Despite their key role in marine and freshwater food webs around the globe, genomic resources for zooplanktonic species are still scarce. In many aquatic food webs, zooplanktonic crustaceans link primary production by

phytoplankton and secondary consumers, such as planktivorous fish and larger invertebrate species (Gannon & Stemberger, 1978; Gliwicz, 1990; Lampert & Sommer, 2007).

Daphnia are highly phenotypically plastic and a textbook example for inducible defense mechanisms (Tollrian & Harvell, 1999), as they respond to variation in predation risk through spectacular changes in morphology. Further, *Daphnia* are cyclical parthenogens and hence able to alternate between asexual and sexual reproduction. They reproduce asexually through longer periods of time, and the product of sexual reproduction events (usually seasonal) are resting eggs able to withstand adverse conditions for decades and even centuries (Frisch et al., 2014). Resting eggs extracted from sediment cores can be hatched, and ancient genotypes brought to life (reviewed in Orsini et al., 2013). Moreover, the DNA preserved in those resting eggs can be directly analyzed with various molecular methods (Cousyn et al., 2001; Dziuba et al., 2020; Lack et al., 2018). Thus, cyclical parthenogenesis, biological archives in lake sediments and high levels of phenotypic plasticity make *Daphnia* a particularly interesting model for evolutionary studies.

The genus *Daphnia* is composed of two subgenera, *Ctenodaphnia* and *Daphnia*, and two groups are delimited within the subgenus *Daphnia*: the *D. pulex* group sensu lato and the *Daphnia longispina* group sensu lato (see Adamowicz et al., 2009). The latter is sometimes also referred to as subgenus *Hyalodaphnia* and includes the *D. longispina* species complex (DLSC) (Petrusek, Hobæk, et al., 2008). The two *Daphnia* groups are highly differentiated and share their most recent common ancestor around 30 Ma (MRCA *D. longispina–D. pulex* group, Cornetti et al., 2019). Members of the genus *Daphnia* show little variation in chromosome number, with most species having ten pairs of chromosomes, except for the *D. pulex* group with $n = 12$ (Beaton & Hebert, 1994; Trentini, 1980). All sequenced and assembled *Daphnia* genomes so far belong either to the *D. pulex* group or the subgenus *Ctenodaphnia*, however no high-quality reference genome of the third major group, the *D. longispina* group (*Hyalodaphnia*) is published.

The prevalence of hybridization in the genus *Daphnia* across taxa and ecosystems and its impact on their evolutionary history has intrigued researchers for decades (Schwenk, 1993; Vergilino et al., 2011; Wolf, 1987). In contrast to many other well-studied hybrid systems (Barton & Hewitt, 1985) with clear defined hybrid zones where species' ranges overlap, the distribution of *Daphnia* species and their hybrids is more of a fragmented nature: they occupy lake and pond ecosystems that vary in their ecological characteristics and hence constitute a mosaic across the landscape. Ecologically differentiated taxa and their hybrids are thus distributed across habitat patches (Harrison, 1986). Within these patches, the possibility to interrogate biological archives also revealed fluctuations in *Daphnia* community composition

over time (Alric et al., 2016; Brede et al., 2009), associated with hybridization events among species in some cases. Variation in hybridization events across time and among habitats has often been observed in correlation with ecological changes, such as eutrophication or global change (Brede et al., 2009; Cordellier et al., 2021; Dziuba et al., 2020; Keller et al., 2008; Rellstab et al., 2011; Spaak et al., 2012).

Members of the DLSC inhabit many large ponds and lakes in central and northern Europe, and three of them have been particularly well studied: *Daphnia galeata*, *Daphnia longispina*, and *Daphnia cucullata* (Petrusek, Hobæk, et al., 2008). These species can coexist, but earlier studies suggest gene flow among them is limited (Spaak et al., 2004). Despite their obviously ancient divergence (Schwenk et al., 2000), DLSC species are still able to form interspecific hybrids, although not all combinations are equally likely to lead to viable and fertile individuals (Schwenk et al., 2001). A mechanism preventing gene flow among species might be their different ecological preferences, for example, regarding trophic level (Spaak et al., 2012), food quality (Seidendorf et al., 2007), and predation pressure (Petrusek, Seda, et al., 2008; Spaak & Hoekstra, 1997).

Up to now, genetic markers available to study hybridization in the DLSC are limited to allozymes (Wolf & Mort, 1986), a few mitochondrial regions (Schwenk, 1993), a dozen microsatellite markers (Brede et al., 2006; Thielsch et al., 2012) and a few further nuclear loci (Billiones et al., 2004; Rusek et al., 2015; Skage et al., 2007). Seminal studies such as Brede et al. (2009) and Limburg and Weider (2002) first made use of microsatellite markers to analyze environmentally driven shifts in allelic frequencies, species, and hybrid composition of the DLSC communities in Lake Constance and Belauer See over time, respectively. Further, a number of studies addressed the spatial distribution of DLSC species/taxa with these markers (Griebel et al., 2016; Ma et al., 2019; Thielsch et al., 2017). These low-resolution markers allowed to identify hybrid individuals and brought evidence for introgression but could not provide the resolution necessary to either assess how pervasive introgression is or how it varies across the genome. Further, it is not clear whether introgression occurs among all three species to the same extent. Given the ubiquitous hybridization among the DLSC taxa, the question also arises why they are still well distinguishable species. Whether the DLSC represents a case of incipient speciation, introgression after secondary contact, speciation reversal, or has reached an intermediate constant equilibrium state, among other possibilities, can only be answered with genome-wide analyses empowered by a high-quality genome assembly.

Here, we present a high-quality assembly for *D. galeata*, thus filling an important gap for *Daphnia* whole-genome studies. Furthermore, to facilitate genome-wide assessments of

divergence across species and of introgression between species, we conducted genome-wide resequencing studies in the DLSC. We analyzed whole-genome sequences of parental species and their interspecific hybrids, both from genotypes obtained in the wild and maintained in laboratories, and from single resting eggs extracted from sediment cores. We provide first insights into the intra- and interspecific genome-wide variability in this species complex and identify regions of high divergence. We reconstructed the phylogenetic relationships in the species complex using whole mitochondrial genomes. Finally, we assess the length of ancestry tracts in different classes of hybrids to characterize introgression patterns. Our study paves the way for long-awaited analyses on the dynamics of introgression in this complex and exploitation of the unique opportunity this group has to offer: a window of more than one hundred years of evolution in action.

Results

Genome assembly

The raw assembly was obtained by combining PacBio long reads (1,679,290, 11.52 Gb) and Illumina short reads (70,310,338, 9.79 Gb after trimming) and using the hybrid assembler RA (https://github.com/lbcb-sci/ra, last accessed January 2021). It originally comprised 1,415 contig sequences covering a total length of 153.6 Megabases (Mb), with an N50 value of 172 kilobases (kb) and a slightly elevated GC content (40.02%, Table M3.3) compared with the values expected for a *Daphnia* species (see Table 3.1). According to an analysis based on coverage and GC content of the contig sequences conducted with blobtools (Laetsch & Blaxter, 2017), a portion of the assembly consisted of non-*Daphnia* contigs, which could then be removed (267 contigs, equaling 22.97 Mb, Figure S3.1C). Consequently, GC content decreased to 38.75%, nearing the values obtained for other *Daphnia* assemblies (see Table 3.1 for an overview). The application of this filter as well as the exclusion of the mitochondrial genome led to a decrease in the number of sequences and the total length of the assembly. Iterative scaffolding led to a decrease in the total number of sequences. This together with a substantial increase in N50 resulted in a highly contiguous assembly, with a total length of 133,304,63 bp, an N50 of 756.7 kb and only 346 sequences, that is, on an average 30 sequences per chromosome. Contiguity statistics for the different assembling steps are given in Table M3.3.

Mapping the filtered Illumina reads with bwa mem (Li, 2013) and PacBio reads with Minimap 2.17 (Li, 2018), resulted in a mapping rate of, respectively, 94.1% and 85.5%. The coverage distribution can be seen in Figure S3.1B.

Table 3.1. Assembly metrics and annotation statistics for the present assembly and two previously published *Daphnia* assemblies. Contiguity statistics of the annotation were calculated excluding tRNAscan results. BUSCO 3.0.2 was executed in protein mode for the different MAKER rounds. Conserved domain arrangements were searched with Pfam scan 1.6 and DOGMA 3.4. Results for BUSCO and DOGMA completeness statistics are given in percent.

Species	***D. galeata***	***D. pulex* (Ye et al., 2017)**	***D. magna* (Lee et al., 2019)**
Strain	M5	PA42	SK
Assembly metrics			
No. scaffolds	346	493	4,192
Largest scaffold (bp)	2,950,711	7,584,612	16,359,456
Total length (bp)	133,304,630	189,550,516	122,937,721
N50 (bp)	756,671	1,160,003	10,124,675
L50 (bp)	48	36	5
GC (%)	38.75	40.39	40.54
No. *N*'s	120,845	4,006,006	82,97,703
No. *N*'s per 100 kb	90.65	2,113.42	6,749.52
Annotation			
Number			
Gene	15,845	18,440	15,721
mRNA	16,774	18,440	15,721
Exon	117,364	128,688	95,203
CDS	119,402	118,916	94,047
Mean			
mRNAs/gene	1.06	1	1
Exons/mRNA	7.00	6.98	6.06
CDSs/mRNA	7.12	6.45	5.98
Median length (bp)			
Gene	2,097	1,919.5	1,586
mRNA	2,142	1,919.5	1,521
Exon	167	162	160
Intron	74		
CDS	152	144	159
Total space (bp)			
Gene	51,689,473	53,936,938	37,505,261
mRNA	51,689,329	53,936,938	36,178,687
Exon	29,314,592	30,208,483	22,336,755
CDS	25,132,876	23,586,918	21,881,778
Single			
Exon mRNA	663	144	1,775
CDS mRNA	710	0	0
BUSCO *N* = 1,066			
C	94.3	94.1	97.0
S	91.7	82.6	95.3
D	2.6	11.5	1.7
F	0.7	3.5	1.7
M	5.0	2.4	1.3
DOGMA *N* = 4,222	93.63	91.43	93.91

According to blobtools results, no contamination could be identified in the final assembly (Figure S3.1A). Remaining scaffolds (12, amounting to a total length 1.79 Mb) with taxonomic assignment other than Arthropoda were kept because coverage and GC are similar to *D. galeata* scaffolds and taxonomic assignment alone might be false positive. Further, the completeness assessment through BUSCO (Simão et al., 2015, Arthropoda set, odb9) indicated 95.7% of complete single copy core orthologs and a very low duplication rate (C: 95.7% [S: 94.7%, D: 1.0%], F: 0.8%, M: 3.5%, n: 1,066). The genome size was estimated based on mapped nucleotides and mode of the coverage distribution by backmap 0.3 (https://github.com/schellt/backmap, last accessed January 2021), resulting in 156.86 and 178.03 Mb for Illumina (52×) and PacBio (26×) respectively, and by k-mer based approach using GenomeScope resulting in a size of 150.6 Mb. When compared with other published full genomes for *Daphnia* species, the *D. galeata* final assembly is shorter than both *D. pulex* assemblies (Colbourne et al., 2011; Ye et al., 2017), and roughly the same size as *Daphnia magna* (Lee et al., 2019), which also has ten chromosomes (Table 3.1). The GC content is lower, which can be attributed to the strict filtering for contamination applied pre- and post-assembly, a procedure not applied in the other *Daphnia* assemblies, to our knowledge. Even though Lee et al. (2019) and Ye et al. (2017) treated the animals with antibiotics before sequencing this suggests that these genome assemblies contain more contigs of bacterial origin than the *D. galeata* assembly. Thanks to the use of long-read data, iterative scaffolding and gap filling; the number and length of assembly gaps (Ns) are substantially lower and contiguity is high (but see Table S3.1).

Genome annotation

After applying RepeatMasker (Smit et al., 2013–2015) with the custom repeat library described in the Materials and Methods section, 21.9% of the assembly was masked. The distribution of masked fraction per repeat element can be found in Table M3.5.

The final annotation with MAKER (Holt & Yandell, 2011) predicts 15,845 genes with a median length of 2,097 base pairs. There is an average of 1.06 mRNAs per gene and seven exons per mRNA (Table 3.1). The total number of predicted mRNA substantially differs from the number of transcripts previously published for this species (32,903, Huylmans et al., 2016). This is not surprising, as this transcriptome assembly did not make use of protein evidence we included here, and might contain isoforms. Further, it was based on a pool of mRNA from different clonal lines, and the assembly process might have been impeded by allelic diversity. As further quality criterion, the Annotation Editing Distance (AED) was compared across the three MAKER rounds and is visualized in Figure M3.4. AED improved mostly between rounds 1 and 2 of the annotation but only marginally with a further round.

A high percentage of protein sequences could be annotated: 15,898 (94.78%) with InterProScan (Jones et al., 2014) and 15,960 (95.15%) with BLAST against Swiss-Prot. With this combination of searches, a hit within InterProScan and BLAST was found for 16,675 protein sequences (99.41%). GeneOntology (GO) annotation was possible for 9,555 sequences (56.96%). A detailed overview of the functional annotated sequences per database or search algorithm is shown in Table M3.7.

Genotyping

Short-read sequence data were generated for 72 individuals: 17 unamplified DNA samples from isofemale clonal lines and 55 whole-genome amplification (WGA) samples (conducted on single resting eggs) that passed PCR contamination checks. After screening for contamination and removing data sets with only very few reads mapping to the *D. galeata* reference, 49 single genotypes remained: 32 from resting eggs and 17 from clonally propagated lines, established from individuals sampled in the water column or hatched from resting eggs (Figure 3.1A and Table S3.2). Data gained from clonal lines with a species attribution were used as “parental species” data: five samples for *D. galeata*, four for *D. longispina*, and three for *D. cucullata*. The parental clones are part of two larger clone panels representing the parental species and their diversity in several European lakes. Their identity was established prior to this study either based on mitochondrial and microsatellite markers (M5, LC3_6, J2, Herrmann et al., 2017) or morphological examination, mitochondrial markers and factorial correspondence analyses based on microsatellite markers (Alric et al., 2016; Möst, 2013), including hybrids and historical resting eggs, which separates parental species and hybrids (Alric et al., 2016; Dlouhá et al., 2010; Rellstab et al., 2011; Yin et al., 2014). In addition, data were available for four resting eggs from Arendsee (AR), 12 resting eggs from Dobersdorfer See (DOB), five clonal lines and eight resting eggs from Eichbaumsee (EIC), and eight resting eggs from Selenter See (SE) (Table S3.2). Although the analysis of eggs from older sediment layers was attempted, biological material was either limited, of poor quality, or contaminated. Our isotope dating for DOB was unconclusive: either slides of the cored location or a high sedimentation rate meant the top 30 cm of the core did not show the usual isotope peaks, thus preventing precise dating. EIC samples were recent since they were collected from surface bank sand. For SE, the oldest eggs analyzed here originated from the 2–3 cm layer of the core, which corresponds to max. ~17 years (Andersen T, personal communication). For AR the oldest eggs for which results were obtained were isolated from the 4–5 cm layer, corresponding to ~2005 (Bálint M, personal communication).

An average of 89.9% (range: 31.7–98.6%) reads aligned to the reference genome with a mean coverage of 10.26× (range: 0.34–52.30×) (Table S3.3). The final SNP data set for subsequent analyses after quality-filtering included 3,240,339 SNPs across the 49 samples. To rule out

possible reference bias, we compared mapping rates of reads with the reference allele and to the alternative allele at heterozygous sites. We found no preferential mapping of the reference allele, as all species categories and the hybrids had a median distribution close to 0.5 (Figure S3.2).

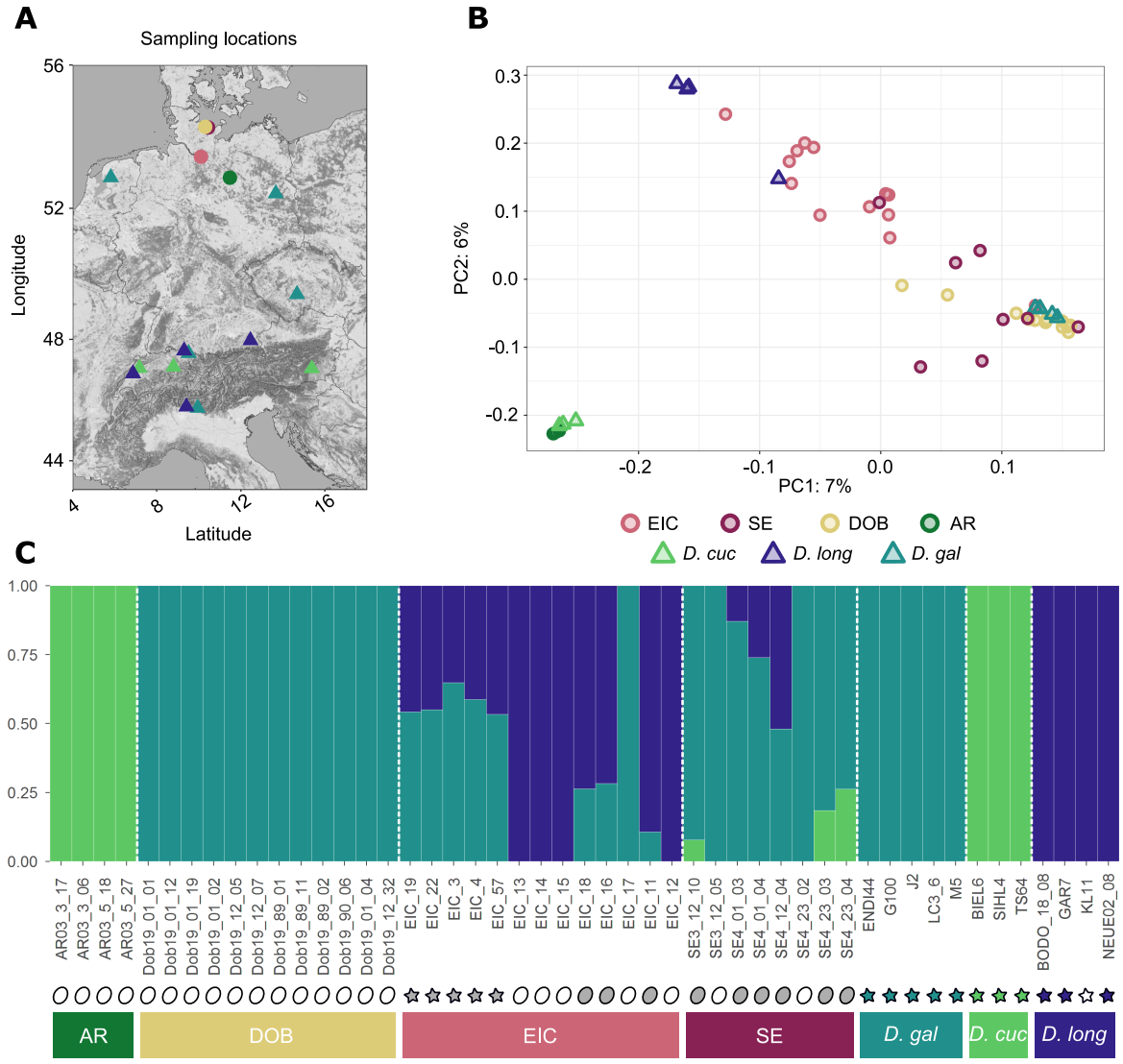


Figure 3.1. Parental species: *D. gal*: *D. galeata*, *D. long*: *D. longispina*, *D. cuc*; *D. cucullata*, populations: AR, Arendsee; Dob, Dobersdorfer See; SE, Selenter See; EIC, Eichbaumsee. Color and symbol coding are consistent throughout panels (A) and (B). (A) Map of the sampling locations. (B) PCA plot obtained with SNPrelate, including loci with linkage r^2 <0.5 within 500-kb sliding windows. (C) Admixture plot obtained with $K = 3$. Symbols indicate the sample type: oval for genotypes sequenced directly from resting eggs, stars for genotypes sampled in the water column and propagated clonally in the laboratory prior to sequencing. Symbol filling indicates how these genotypes were classified in subsequent analyses: white for nonadmixed genotypes; gray for admixed genotypes; and green, blue, and teal for genotypes used as representatives for parental species. Bottom bars are color coded to match the color scheme used in panels (A) and (B).

Principal component analysis

In a PCA including all genotypes and conducted with SNPRelate v1.20.1 (Zheng et al., 2012), the parental species genotypes grouped in three very distinct clusters. *Daphnia cucullata* separated from *D. galeata* along PC1, which explained 7% of the variation. *Daphnia longispina*

separated from *D. galeata* and *D. cucullata* along PC2 which explained 6% of the variation (Figure 3.1B). Although sampled in different lakes, all parental species genotypes were grouped in tight clusters along the two axes with little evidence for population substructure. Population AR clustered with the *D. cucullata* reference individuals while population samples from DOB, EIC, and SE were more spread out, mostly between the *D. galeata* and *D. longispina* clusters.

Admixture analyses uncover hybrids

The PCA results are confirmed by an admixture analysis conducted with ADMIXTURE (Alexander & Lange, 2011) with $K = 3$, supported by the lowest cross-validation error of the tested *K* values. The known parental species genotypes were clearly separated into three clusters (Figure 3.1C). Although we detect no evidence of admixture in the AR and DOB samples and, based on our parental species, consider them to belong to the species *D. cucullata* and *D. galeata*, respectively, the two other populations seem to consist mostly of admixed individuals. The five EIC samples sequenced after clonal propagation were all found to be admixed (*D. galeata* and *D. longispina*), whereas the EIC resting eggs were either admixed (3) or belonged to one of the parental species (5). SE resting eggs present all combinations of admixture except *D. cucullata*×*D. longispina*: *D. galeata* (2), admixed between *D. galeata* and *D. cucullata* (3), and admixed between *D. galeata* and *D. longispina* (3).

Ancestry painting

Based on results obtained in the ADMIXTURE analysis, two pairs of species and their putative hybrids were analyzed with an "ancestry painting" approach, outlined in Barth et al. (2020) and Runemark et al. (2018): *D. galeata* and *D. longispina* parental genotypes and putative hybrids between them from populations EIC and SE, and *D. galeata* and *D. cucullata* parental genotypes and putative hybrids between them from population SE. Briefly, after identifying fixed sites for each of the species in the analyzed pair, heterozygosity was calculated for these sites and a hybrid index derived from the obtained results (https://github.com/mmatschiner/tutorials/tree/master/analysis_of_introgression_with_snp_data, last accessed November 2021). Further, information on the maternal species is used to tentatively categorize the admixed individuals. For a first-generation hybrid (F1) the expectation would be 50% of the nuclear genome being derived from each parental species (hybrid index ≈ 0.5) and mostly heterozygous fixed sites (heterozygosity ≈ 1.0). Individuals originating from the backcrossing of F1 with one of the parental species are expected to have hybrid index values around 0.25 or 0.75 (Figure 3.2D). We consider individuals with intermediate hybrid indices (>0.25 and <0.75) and lower heterozygosity (<0.5) to be later-generation hybrids, meaning they have one or multiple hybrid ancestors we are not able to classify further (Slager et al., 2020). We

consider individuals with a hybrid index of ≤0.25 or ≥0.75 to be backcrossed with the respective parental species in at least one generation and the majority of the genome derives from one species. This definition is broad and will be refined with the addition of a greater number of parental genotypes.

The comparison of genotypes from parental species *D. galeata* and *D. longispina* (five and three individuals, respectively) allowed identifying a total of 335,052 fixed sites between the two species. Due to the quality filters applied to parental and hybrid genotypes, we could analyze 131,914 of these fixed sites in the putative hybrids, where read coverage was sufficient. The diploid genotypes were then plotted for all hybrids as homozygous for either of the parental species or as heterozygous (Figure 3.2A for the 50 longest scaffolds). The *D. longispina* reference clone KL11 was excluded from further analysis due to issues with missing data. All 11 genotypes from SE and EIC identified as likely *D. galeata*×*D. longispina* hybrids in the ADMIXTURE analysis possessed a *D. galeata* mitochondrial genome. The proportion of the maternal *D. galeata* genome in these hybrids, however, varied greatly, between 27.4% and 86.6%, and they all showed low heterozygosity, between 9.1% and 34.6% (Figure 3.2C and Table 3.2). These values are unlikely for F1 hybrids or backcrosses of F1 with one of the parental species (Table 3.2).

Comparing genotypes from the parental species *D. galeata* and *D. cucullata* (five and three individuals, respectively) led to identifying 715,438 fixed sites between the two species (due to quality filtering, 275,216 of these sites were further analyzed). All three *D. galeata*×*D. cucullata* hybrids carried a *D. cucullata* mitochondrial genome, their hybrid index varied between 0.079 and 0.267 and their heterozygosity ranged from 6.4% to 50.9% (Figure 3.2B and C; Table 3.2). The individual SE_23_04 is most likely the result of a backcrossing with *D. galeata*; however, it is difficult to determine what backcrossed with it: either an F1 hybrid or a later-generation hybrid, that is, that resulted from several generations of admixture. Haplotype information would be needed to gain certainty. The other two hybrids' lower hybrid index hints at backcrossing with *D. galeata*, according to the criteria defined above.

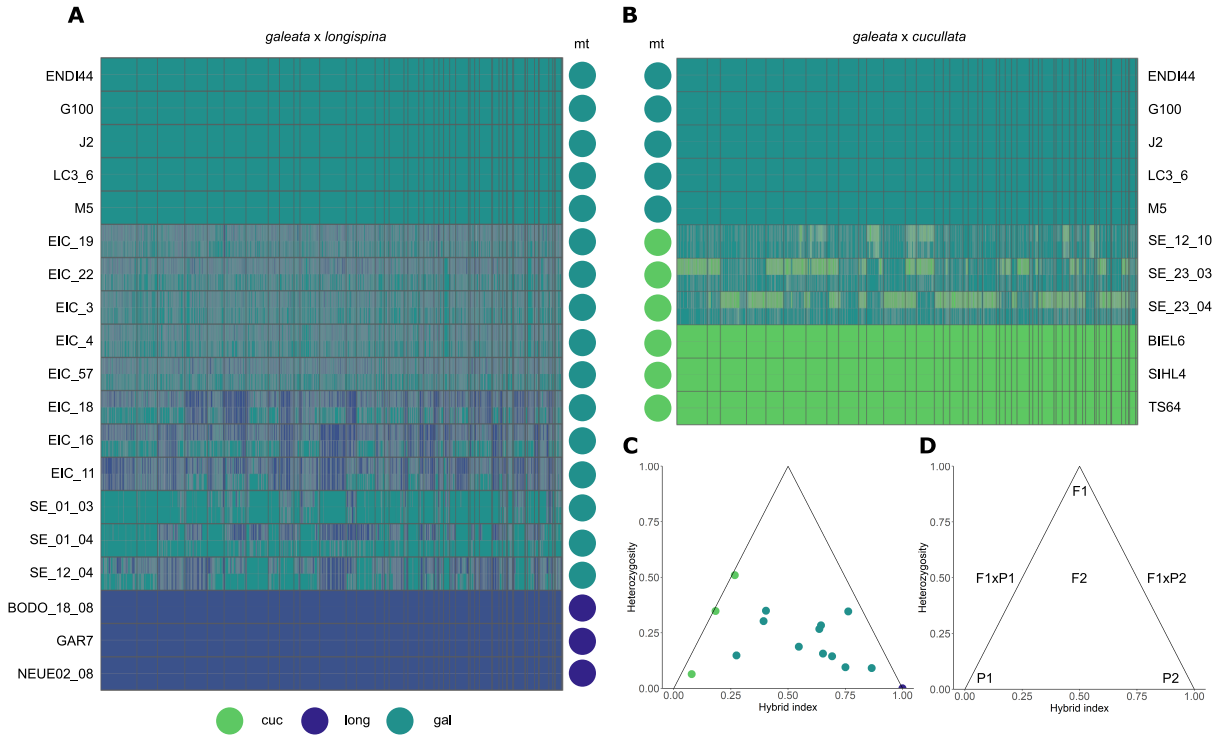


Figure 3.2. (A, B) Ancestry painting of the hybrid individuals identified through the admixture analysis. Each row represents an individual. Colored circles on the side indicate the mitochondrial identity of the individuals, based on the analysis of full mitochondrial genomes. Scaffolds are sorted by length and separated by thin gray lines. In panels (A) and (B), the five upper rows represent individuals assigned to the parental species *Daphnia galeata*. In (A), the last three rows correspond to individuals assigned to the parental species *D. longispina*. In (B), the last three rows correspond to individuals assigned to the parental species *D. cucullata*. Triangle plots summarizing (C) the hybrid index and mitochondrial species identity for all individuals identified as admixed, and (D) The hypothetical expected means of parental species (P1 and P2) and hybrid classes (F1xP1 and F1xP2: backcrosses with parental species P1 and P2, respectively).

Table 3.2. Data derived from ancestry painting analysis and based on the fixed sites inferred from analyzing parental species genotypes. Maternal species attribution is based on mitochondrial phylogeny, hybrid attribution is based on the ADMIXTURE plot.

Sample	Hybrid Index	Heterozygosity	Maternal Species	Hybrid	Interpretation
SE_12_10	0.079	0.064	cuc	gal×cuc	Backcross gal
SE_23_03	0.183	0.348	cuc	gal×cuc	Backcross gal
SE_23_04	0.267	0.509	cuc	gal×cuc	Unclear
EIC_19	0.653	0.156	gal	gal×long	Later-generation
EIC_22	0.644	0.283	gal	gal×long	Later-generation
EIC_3	0.751	0.095	gal	gal×long	Backcross gal
EIC_4	0.693	0.144	gal	gal×long	Later-generation
EIC_57	0.637	0.267	gal	gal×long	Later-generation
EIC_18	0.393	0.302	gal	gal×long	Later-generation
EIC_16	0.403	0.349	gal	gal×long	Later-generation
EIC_11	0.274	0.148	gal	gal×long	Later-generation
SE_01_03	0.866	0.091	gal	gal×long	Backcross gal
SE_01_04	0.763	0.346	gal	gal×long	Backcross gal
SE_12_04	0.547	0.187	gal	gal×long	Later-generation

Population genomics parameters

To calculate genome-wide nucleotide diversity (π), between-taxon differentiation (F_{ST}), and between-taxon divergence (d_{xy}) within 100-kb sliding windows, we took advantage of the inference made with ADMIXTURE and pooled all genotypes which were unambiguously assigned to either of the parental species clusters. Consequently, a total of seven genotypes from four populations were classified as *D. cucullata*, eight from five populations as *D. longispina*, and 20 genotypes from eight populations as *D. galeata* (Table S3.2). All values (d_{xy}, π, and F_{ST}) were calculated with the script popgenWindows.py (github.com/simonhmartin/genomics_general, release 0.3, last accessed February 2021) for each species pair and are plotted for the 50 largest scaffolds in Figure 3.3.

The window-based F_{ST} values for all three possible pairs among the three species averaged 0.274 for *D. galeata* versus *D. longispina*, 0.343 for *D. cucullata* versus *D. longispina* and 0.364 for *D. galeata* versus *D. cucullata*. The mean sequence divergence d_{xy} for the three pairs was 0.018 for *D. galeata* versus *D. longispina* and 0.022 for both *D. cucullata* versus *D. longispina* and *D. cucullata* versus *D. galeata*. Both parameters show similar patterns, with lower values on an average when comparing *D. galeata* to *D. longispina* than when comparing *cucullata* to either one of the other species. These patterns confirm the results obtained with other analyses, for example, the higher number of fixed sites between *D. galeata* and *D. cucullata* in the ancestry painting analysis.

The window-based estimates show high variability in levels of differentiation and divergence along the genome. Further, regions of high or low differentiation are mostly associated with depleted or high nucleotide diversity, respectively (see scaffolds 2 and 9, for example). However, the genome being represented by unordered scaffolds instead of chromosomes makes this difficult to interpret further.

Nucleotide diversity (π) to quantify the level of genetic variation within each taxon was on average higher for *D. longispina* (1.18%) than for the other two species (0.95% and 0.85% for *D. galeata* and *D. cucullata*, respectively). This cannot be explained by the differences in group sample sizes, since *D. galeata* was the group with the largest sample size (and highest number of sampled populations). To ensure our window-based estimates were not biased because of the overrepresentation of some populations in a group (e.g., DOB in the *galeata* group), we also calculated these indices using only one individual from each population per species; if one population contained multiple individuals, we picked one individual at random to represent this population (see Table S3.2, for a listing of the used genotypes—results shown in Figure S3.4).

Many more highly differentiated windows and genes were shared among two or all species pairs than would be expected by a random intersection (Figure S3.5). For example, a total of

2,575 10-kb windows had an F_{ST} value within the 95th percentile in the pair *D. galeata*/*D. longispina* and 2,569 in the pair *D. galeata*/*D. cucullata*. The mean expected number of windows in common between these two pairwise comparisons was 113, but the number of windows in common observed in the data was 1,601. A similar pattern was observed in all other intersections. This result suggests that the location of differentiated genome parts is not due to random processes but has biological significance. A GO-enrichment analysis of these isolated genes to shed light on the function of these species-specific genes, however, was not possible, because of the low number of genes with GO annotation. For the pair *D. galeata*/*cucullata*, only 12% of the genes in the outlier windows were annotated with Gene Ontologies, for the pair *D. galeata*/*longispina* it was 11% and for the *D. cucullata*/*longispina* pair it was 10%.

Phylogeny based on complete mitochondrial genomes

Phylogenetic reconstruction based on the mitochondrial protein-coding and ribosomal RNA (rRNA) genes were largely consistent with earlier mitochondrial phylogenies based on single or few mitochondrial genes (Adamowicz et al., 2009; Petrusek et al., 2012). We identified highly supported clades comprising the respective parental genotypes, hence representing *D. longispina*, *D. cucullata*, and *D. galeata* mitochondrial haplotypes (Figure 3.4). *Daphnia cucullata* and *D. galeata* mitochondrial haplotypes clustered as sister groups. Although the mitochondrial haplotypes in the *D. longispina* and *D. cucullata* clusters do not show much divergence, the *D. galeata* haplotype cluster also contains deeper branching events (haplotype EIC_15 and AR3_17/AR5_18). Further, although all samples from AR were unequivocally categorized as *D. cucullata* in the ADMIXTURE analysis and clustered with *D. cucullata* parental genotypes in the PCA, two of them have mitochondrial haplotypes falling into the *D. galeata* cluster (AR3_17 and AR5_18). A similar mismatch was also observed for EIC_15, which clusters with *D. longispina* when considering nuclear SNP and with *D. galeata* when considering the mitochondrial genome. Within the species clusters, we observed a grouping by lake with many haplotypes being either identical or very similar when originating from the same location. The trees obtained with either only protein-coding genes (PCGs) (CODON model) or PCGs and rRNA genes but with a mixed model (DNA for rRNAs and CODON for DNA) were all consistent with the tree shown here and are therefore not included.

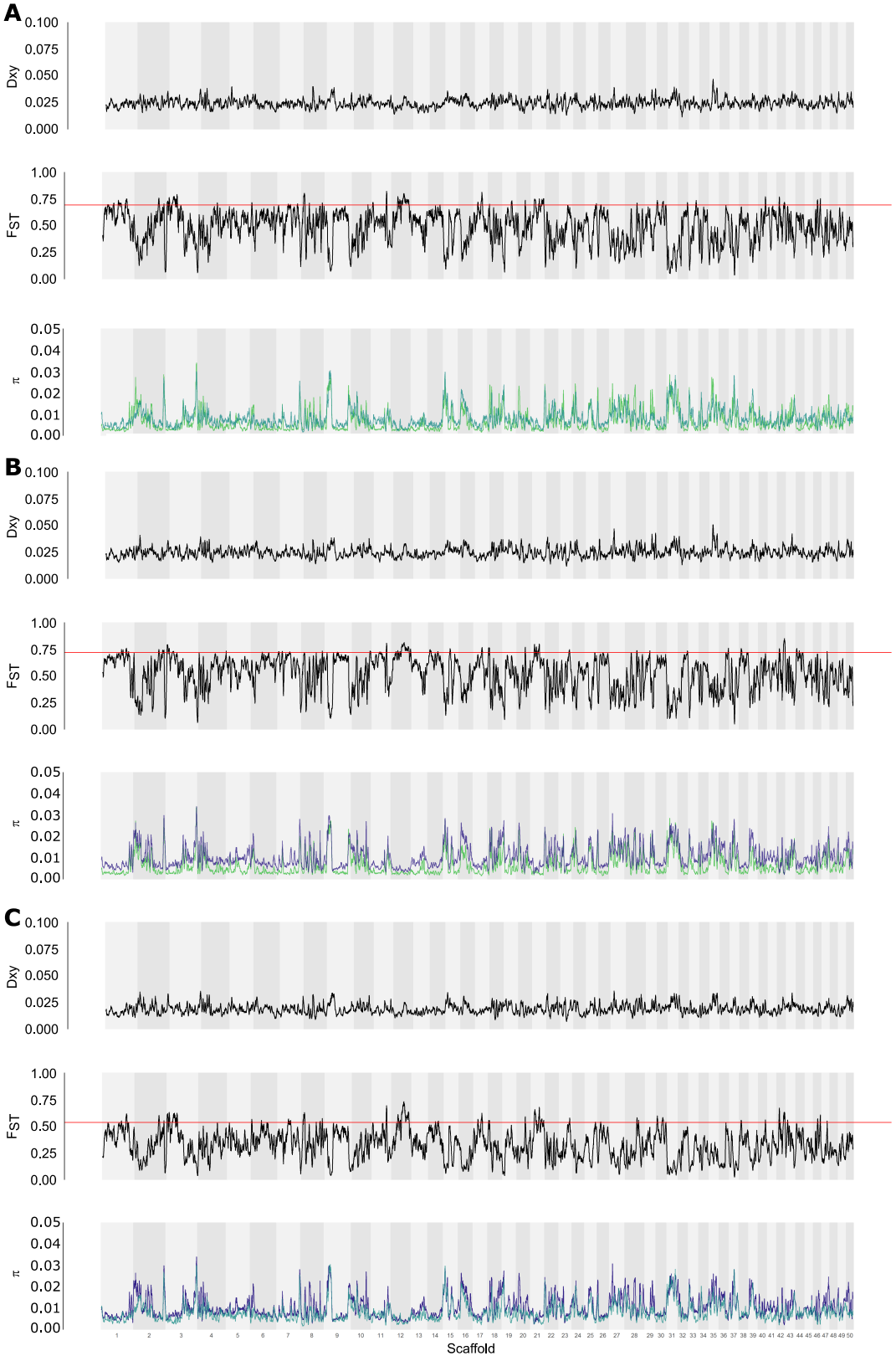


Figure 3.3. Window-based statistics for the pairs (A) *Daphnia galeata*/*D. cucullata*, (B) *D. cucullata*/*D. longispina*, and (C) *D. galeata*/*D. longispina*, shown for the 50 largest scaffolds in 100-kb windows with 10-kb step size—calculations are for all nonadmixed individuals unambiguously assigned to parental species according to the ADMIXTURE analysis. In each panel from top to bottom: d_{xy} values, pairwise F_{ST} values with a red horizontal line indicating the 95th percentile, nucleotide diversity (π) for *D. galeata* (teal), *D. longispina* (dark blue), and *D. cucullata* (lime green).

Patterns of introgression

We tested all four northern Germany populations (EIC, DOB, AR, and SE) for admixture between the three reference species with f_3 statistics tests (Table 3.3) and considered a *Z*-score <−3 as significant (following Patterson et al., 2012; Reich et al., 2009). Negative and significant values ($f_3 = -0.19$) using EIC as the test population and *D. galeata* and *D. longispina* as the source populations indicated mixed ancestry from these two or closely related populations. For population SE, the f_3 test supports both admixed ancestry from *D. galeata* and *D. longispina* ($f_3 = -0.09$) and *D. galeata* and *D. cucullata* ($f_3 = -0.15$). All tests for population DOB and AR were positive providing no evidence of admixture events. The supported introgression events are consistent with the results in our previous analyses conducted with ADMIXTURE and the ancestry painting approach.

Table 3.3. Summary of the f_3 Statistic for Admixture in the Form (C; A, B). A significantly (*Z*-score <−3, in bold) negative f_3 value implies that the target population C is admixed. SE, standard error.

Source population A	Source population B	Target population C	f_3	SE	*Z*-score	No. Sites
gal	long	AR	5.48226	0.154392	35.509	1,072,106
gal	cuc	AR	0.245124	0.006208	39.487	791,570
long	cuc	AR	0.225393	0.005475	41.165	875,293
gal	long	Dob	0.152725	0.003233	47.242	752,265
gal	cuc	Dob	0.150027	0.003028	49.546	815,035
long	cuc	Dob	1.763147	0.049657	35.506	1,059,418
gal	long	EIC	−0.193114	0.003224	**−59.898**	864,328
gal	cuc	EIC	0.086614	0.002623	33.025	1,119,715
long	cuc	EIC	0.014322	0.001774	8.071	1,145,530
gal	long	SE	−0.093233	0.003222	**−28.939**	1,029,744
gal	cuc	SE	−0.154696	0.003924	**−39.426**	955,516
long	cuc	SE	0.288212	0.011203	25.727	1,092,186

We performed local ancestry inference with Loter (Dias-Alves et al., 2018) to trace genome-wide introgression among the hybrids and infer additional details about the parental species and backcross history from haplotype information. The results were summarized genome-wide for the ancestry proportion, heterozygosity of ancestry and the number of ancestry transitions where each ancestry tract is counted when the state of an SNP changes to the other species or at the end of a scaffold (Figure 3.5). The three *D. galeata*×*cucullata* hybrids were all found to have high *galeata* ancestry (73.9–94.2%) and the individuals SE_23_03 and SE_23_04 were confirmed as the offspring of a later-generation hybrid and a pure *galeata* parent (Figure S3.6B).

Seven *D. galeata*×*longispina* hybrids had very high *galeata* ancestry (81.4–97.9%), and visual inspection of the ancestry tracts (Figure S3.6A) revealed very short *longispina* tracts and scaffolds with multiple breakpoints indicating multiple generations of recombination. Four *D.*

galeata×*longispina* hybrids had lower *galeata* ancestry (27.7–59.0%) and the presence of complete *longispina* scaffolds, implying some backcrossing with *longispina*. The haplotype phasing confirmed that the parents of all *D. galeata*×*longispina* hybrids were also of hybrid origin. The average and maximum ancestry tract length for all *D. galeata*×*longispina* hybrids is shorter than those for *D. galeata*×*cucullata* hybrids.

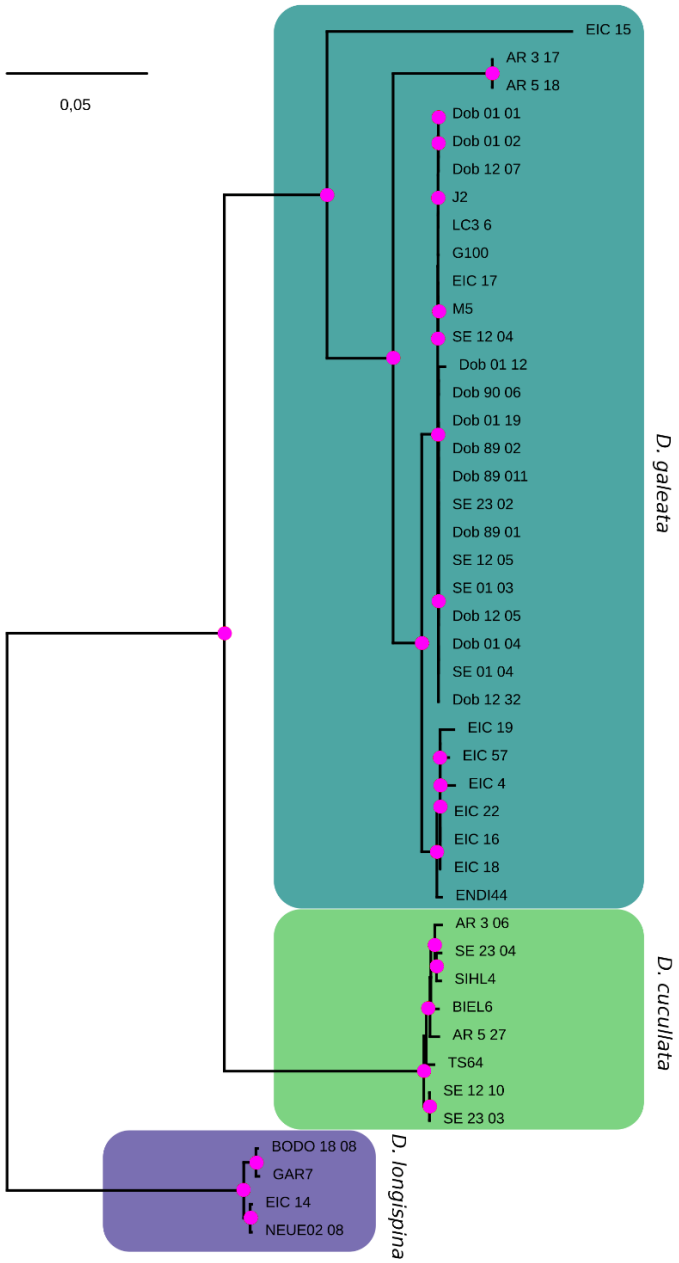


Figure 3.4. Maximum-likelihood tree reconstructed from mitochondrial PCGs and rRNA genes of parental species, clones sampled in the water column and resting eggs sequenced in this study. The tree reveals distinct and highly supported clusters corresponding to *Daphnia galeata*, *D. cucullata*, and *D. longispina* mitotypes (as defined by the respective parental species and a sister taxa relationship between *D. galeata* and *D. cucullata)*. Here, the best tree (log*L* = −47,950.82) rooted with outgroup *D. laevis* is depicted. Magenta dots indicate Shimodaira–Hasegawa approximate likelihood ratio test values ≥80% and ultrafast bootstrap support values ≥95% calculated from 10,000 bootstrap replicates (SH-aLRT/ UFboot). The scale bar corresponds to 0.05 nucleotide substitutions per nucleotide site. KL11 was excluded due to missing data.

Discussion

A reference genome for studying a species complex

Daphnia are a key species in freshwater habitats. Previous studies have established reference genomes for the model species *D. pulex* (Colbourne et al., 2011; Ye et al., 2017) and *D. magna* (Lee et al., 2019). No high-quality reference genome for species belonging to the DLSC was available so far. To date, it is unclear whether the ecological differentiation and/or intrinsic incompatibilities drive and maintain divergence between DLSC species. Besides its utility for studies of hybridization events in the DLSC, the new assembly we present here will thus allow us to better understand the evolution of a key species in European freshwaters.

Even though the onset of the DLSC radiation was dated to 5–7 Ma based on nuclear and mitochondrial markers (Adamowicz et al., 2009; Schwenk, 1993; Taylor et al., 1996), several factors confirm the suitability of this reference for all tested species. Mapping success and coverage of whole-genome data from *D. cucullata* and *D. longispina* to the reference genome were high, and we found no evidence of reference bias. This assembly clearly benefited from advances both in the sequencing technologies and assembly and postprocessing algorithms since the first *Daphnia* genome (Colbourne et al., 2011). The metrics used for assessing its quality reveal that in particular, the combination of long- and short-read technologies led to highly contiguous and accurate scaffolds. Although we likely did not recover the genome in its full length (133 Mb out of an estimated 156 Mb), and the N50 value is lower than those obtained for *D. pulex* (Ye et al., 2017) and *D. magna* (Lee et al., 2019), iterative scaffolding allowed for a very efficient gap-closing, and an exceptionally low number of mismatches, compared with the other *Daphnia* assemblies.

Pervasive introgression in the DLSC

We utilize a method that allows us to interrogate biological archives and analyze whole *Daphnia* genomes directly from the resting egg bank (Lack et al., 2018) without hatching and culturing several clonal lineages. This provides a wide sweep of populations, past and present, with each egg being the product of local sexual recombination.

Although no evidence of introgression was found in the DOB population, the three other locations host a variety of admixed genotypes. SE and EIC can even be considered hybridization hotspots with more than 60% of individuals having hybrid ancestry, as revealed in the ADMIXTURE analysis. However, Kong and Kubatko (2021) very recently showed that ADMIXTURE is sensitive to unequal contributions by the parental species, and we thus sought to support these inferences by f_3 calculations and using an ancestry painting approach.

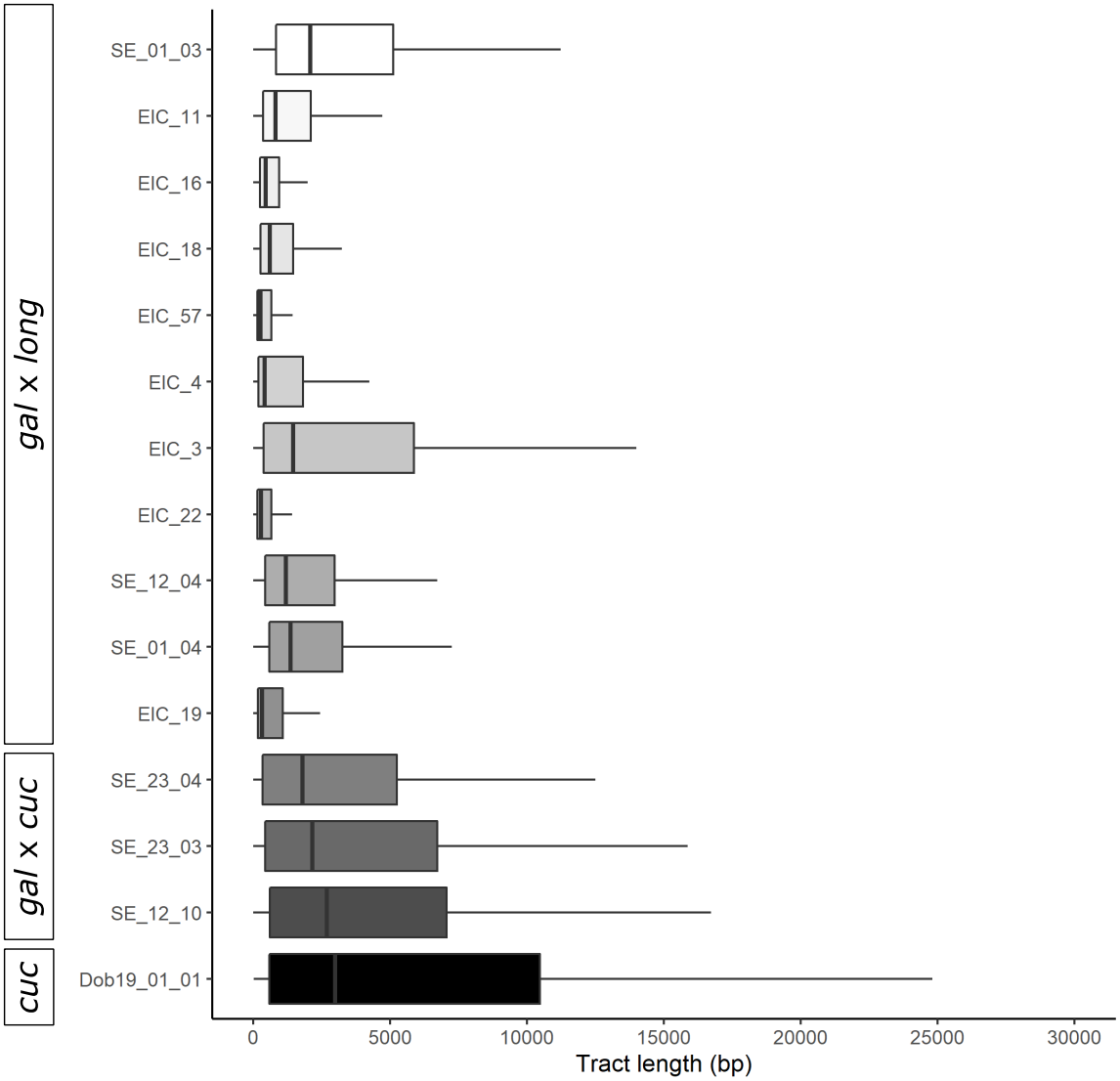


Figure 3.5. Distribution of the ancestry tract length where each ancestry tract represents the state of a SNP changing to the other species or the end of a scaffold in the local ancestry inference for each admixed individual and one nonadmixed *D. cucullata* individual. The nonadmixed individual displays the ancestry tract length distribution when all scaffolds derive from the same species. Hybrid type (according to ADMIXTURE analysis) is given on the left side.

In DOB, ADMIXTURE delivered unequivocal results. Further, the f_3 index indicated that no introgression was detectable in this population. However, the PCA plot shows that some of the DOB genotypes are near hybrid individuals. ANGSD results are similar but these genotypes are nearer to the parental species (Figure S3.3A). To address these slightly contradictory results, we therefore conducted an ancestry painting on two DOB genotypes, (12_07 and 89_02, Table S3.7). Both genotypes had very low heterozygosity, thus confirming the ADMIXTURE and f_3 outcomes. A possible explanation would be that these two genotypes carry variation that is not reflected in our limited sampling of the parental species. When comparing with microsatellite-based analysis including many more populations and data points (Thielsch et al., 2009), the *D. galeata* cluster has often been larger and more diverse than the others. The seemingly two “stray” DOB genotypes are therefore likely well within the species variation

boundaries. All mitochondrial haplotypes were clustered together in the phylogenetic reconstruction as well.

In AR, despite the high resting egg density found in the sediment, only very few could be successfully genotyped. Although all inferences based on nuclear markers (PCA, ADMIXTURE, f_3) indicated an absence of hybridization or introgression in this population, the mitochondrial phylogenetic reconstruction showed diverging results. From a nuclear point of view, all genotypes could be categorized as *D. cucullata*, but two out of four AR individuals presented the mitochondrial genome of another species, that is, *D. galeata*. However, the phylogenetic reconstruction shows that the two AR mitochondrial haplotypes form a cluster separate from the main *D. galeata* cluster, which hints at different evolutionary history for these mitochondrial genomes. Such distinct lineages within a species and mito-nuclear discordances were also found by Thielsch et al. (2017) in the DLSC and mitochondrial capture has been detected in other *Daphnia* species (Marková et al., 2013). It is an interesting phenomenon in the DLSC that merits to be further investigated in the future with broader sampling.

In EIC, both pelagic samples and resting eggs were analyzed. Genotypes sampled alive from the water column were all inferred to be admixed to various degrees, three resting egg samples were also admixed, and the remaining five were assigned to either one of the two parental species. We conducted the ancestry painting approach on all EIC individuals; the fixed sites heterozygosity of the individuals categorized as nonadmixed in ADMIXTURE was indeed near zero (Table S3.7). Such high abundances of *D. galeata×longispina* hybrid resting eggs in periods of rapidly changing environmental conditions (i.e., eutrophication) have also been recorded in Lake Constance (Brede et al., 2009). The high frequency of *D. galeata×longispina* hybrids observed here might be due to similar ecological history: the lake Eichbaumsee was created through sand excavation for construction work around ~40 years ago and is characterized by extreme eutrophication and even hypertrophy that could not be remediated. The presence of later-generation hybrids and backcrosses with *D. galeata* and *D. longispina* and short ancestry tract length suggest that both species have been present and hybridizing for most of the lake's short history, or even that it was colonized by individuals of hybrid origin. However, we only obtained contemporary samples for EIC and analysis of resting eggs from sediment cores are needed to distinguish between the two hypotheses.

In SE, diversity is high, both in terms of species combinations in admixed individuals and in terms of degrees of introgression. Although we analyzed eggs from sediments cores, they originated from the first centimeters and there is, therefore, no clear temporal pattern that separates the different hybrid combinations found here. Strikingly, although SE and DOB are geographically very close to each other (~10 km), and dispersal of resting eggs through, for

example, waterfowl or storms would be possible (Figuerola et al., 2005; Frisch et al., 2007; Pietrzak & Ślusarczyk, 2006), the *Daphnia* communities are quite different. This might be due to their different eutrophication levels, reflecting the fact that initial colonization was followed by the establishment of different species, or a combination of both. The observed diversity at such a small spatial scale underlines the mosaic nature of freshwater habitats and the usefulness of approaches including many populations to fully understand genetic diversity arising from colonization and hybridization events in the DLSC.

Previous studies using mitochondrial and few nuclear markers (Alric et al., 2016; Thielsch et al., 2012) were able to categorize hybrids into F1, F2, and backcrosses. However, due to the low resolution of the used markers, further categorizing and above all identification of genome-wide breaking points was not possible at the time. The *D. galeata* reference genome and resequencing data offer now a much higher resolution to assess later-generation hybrids and patterns across the genome. In general, hybrids identified in this study seem to have a complex history of multiple generations of hybridization and backcrossing with both parental species that we are not able to detangle using only ancestry paintings. The local ancestry inference revealed that the average ancestry tract length for *D. galeata*×*longispina* hybrids from EIC and SE is shorter than those for *D. galeata*×*cucullata* hybrids. There are several explanations for the observed pattern.

One is that more generations of recombination led to shorter introgressed tracts, and the *D. galeata*×*longispina* hybrids are therefore the result of a greater number of sexual generations than the *D. galeata*×*cucullata* hybrids. The genomic mosaic of ancestry segments for all hybrid individuals is also characterized by multiple breakpoints within the same scaffolds, which is only possible after multiple generations of recombination. However, data on genome-wide recombination rates and selection are needed to reach solid conclusions about the correlation between tract length and age of the hybridization event in the individual's ancestors. Alternatively, reproductive isolation might be lower between *D. galeata* and *D. longispina* than between *D. galeata* and *D. cucullata*, thus leading to faster introgression in the former case.

As evidenced by the comparison of genomic windows of higher divergence between species pairs, the introgression pattern is not random: a given region exhibiting high F_{ST} values between the *D. galeata* and *D. longispina* genotypes is also likely to show similarly high F_{ST} values in the *D. galeata*/*D. cucullata* pair. Further, some parts of the genome seem to be effectively shielded from introgression. About a quarter of all genes (4,136) are in regions that are highly differentiated between at least two species and about 5% (859) in parts of the genome that are isolated among all three species of the complex. This is much more than expected by chance (Figure S3.5) and is thus likely due to selection against introgression. It

seems plausible to search among these for genes that conserve the specific identity of the involved taxa, despite incomplete reproductive isolation. Genes responsible for the observed ecological divergence among the taxa (Schwenk et al., 2000) or genetic incompatibilities are most likely candidates to be found in the observed divergent regions. Given the ancient divergence, the speciation process in the DLSC might have attained a selection–migration–drift equilibrium, for which there is growing empirical evidence in other species like stick insects (Riesch et al., 2017), flycatchers (Burri et al., 2015), and nonbiting midges (Schreiber & Pfenninger, 2021). However, the current snapshot could equally likely be a consequence of one or several pulses of hybridization. To assess the stability of the equilibrium, data showing that the introgression/selection process is ongoing and constant across an extended period of time would be required and *Daphnia* offers the unique opportunity to go back in time to test these alternative hypotheses.

New evidence for cytonuclear discordance

The genome-wide perspective also elucidated discordance between nuclear and mitochondrial patterns. The phylogeny based on mitochondrial genomes conforms to previously inferred relationships in the DLSC and suggests *D. galeata* and *D. cucullata* are sister species, with *D. longispina* as an outgroup (Adamowicz et al., 2009; Petrusek et al., 2012). However, several of our analyses based on nuclear SNPs challenge this view and suggest different evolutionary histories for mitochondrial and nuclear genomes. The ancestry painting approach relies on the identification of fixed sites for species in a pairwise comparison. More sites were found to be fixed between *D. galeata* and *D. cucullata* (715,438) than between *D. galeata* and *D. longispina* (335,052), which implies a greater divergence between members of the former pair. Further, F_{ST} values were on average higher between *D. galeata* and *D. longispina* (0.274) than between *D. galeata* and *D. cucullata* (0.364). Reports of cytonuclear discordance are common both in plants (Folk et al., 2017; Huang et al., 2014; Lee-Yaw et al., 2019; Stephens et al., 2015) and animals (Llopart et al., 2014; Melo-Ferreira et al., 2014; Sarver et al., 2021). Several processes can lead to this discordance among closely related species: incomplete lineage sorting causing phylogenetic reconstructions based on mitochondrial markers to differ from the true phylogeny of the taxa, or selection causing the fixation of different mitochondrial genomes in different places from standing variation within species (Barrett & Schluter, 2008). Alternatively, cytonuclear discordance may reflect hybridization between species and cytoplasmic introgression, accompanied or not by selection (reviewed in Sloan et al., 2017). The latter explanation would be quite conceivable in the DLSC.

Conclusion

We here provide the first high-quality resources to study genome-wide patterns of divergence in the DLSC, an ecologically important taxon in European freshwater habitats. By quantifying intra- and interspecific diversity, we provide a first glimpse into introgressive hybridization and lay the ground for further studies aiming at understanding how species boundaries are maintained in the face of gene flow.

Unlike for *D. pulex* and *D. magna*, no linkage groups are known for any species of the DLSC. Hi-C sequencing data will be added in the future to order scaffolds into larger, potentially chromosome-scale scaffolds. Such an approach holds promise in a species complex where laboratory crossings for F2 panels and traditional mapping are nearly impossible. This will allow discovering structural variants, identifying recombination breakpoints along each chromosome and thus provide a deeper understanding of the introgression patterns observed here. The functional role of genes in the regions of high divergence uncovered through this first analysis is yet unclear and will be addressed in future studies.

Finally, wider sampling, with the inclusion of more populations as well as more members of the species complex, and the reconstruction of a nuclear based phylogeny are necessary to reach conclusions about the species relationships and eventually identify the causes of the pattern uncovered here.

Materials and Methods

Sampling

The clonal line used for genome sequencing and assembly, M5, was hatched from a resting egg isolated from the upper layers (first 5 cm, corresponding to the years 2000–2010) of a sediment core taken in Lake Müggelsee in 2010. Further, single genotypes representing the parental species from various locations were used in this study, henceforth “parental species genotypes”. Most of them were established from individuals sampled from the water column and are still maintained through asexual reproduction as monoclonal cultures in the laboratory. Thus, all individuals of a clonal line are the same genotype and can be pooled to achieve sufficient amounts of genomic DNA. The species identity for these genotypes was established through a combination of methods: morphology, mitochondrial sequences, and nuclear markers. Sediment cores were collected from Dobersdorfer See (DOB), Selenter See (SE), and Arendsee (AR), Germany using a gravity corer (Uwitec, Mondsee, AT) (Table S3.4). Samples were taken from the deepest part of the lakes to minimize past disturbance of the sediment. Cores were cut horizontally into 1 cm layers and the layers were stored at 4 °C in

the dark to prevent hatching. Sediment rate of the three lakes was determined using radioisotope dating (^{137}Cs and ^{210}Pb).

In addition, lake sediment from the shoreline of Eichbaumsee (EIC) (Table S3.4), Germany was collected by hand and stored at 4 °C. The exact age of the sediment is unknown but the upper layers most likely contain recent eggs from the last few years. Zooplankton samples were taken from Eichbaumsee using a plankton net (mesh size 150 µm) from which six *Daphnia* clonal lines were established in a laboratory setting with artificial medium (Aachener Daphnien Medium, ADaM; Klüttgen et al., 1994).

All sampling locations are plotted in Figure 3.1A and information on all samples is provided in Table S3.2.

Genome sequencing

DNA extraction for genome sequencing with Illumina and PacBio

DNA was extracted from around 60 clonal M5 individuals collected from batch cultures maintained in ADaM, and fed with the algae *Acutodesmus obliquus*, cultivated in medium modified after Zehnder and Gorham (1960). Extraction was conducted following a phenol chloroform-based protocol with an RNase step and subsequently sequenced on an Illumina HiSeq4000 at BGI China. Additionally, tissue samples with around 3,000 individuals were sent to BGI for DNA extraction and PacBio sequencing.

Resequencing (Population genomics approach)

DNA extraction from batch cultures for resequencing

For clonal lines used as reference for the parental species, individuals were raised in batch cultures and treated with antibiotics prior to collection and storage at −20 or −80 °C. DNA was extracted with either a phenol chloroform method, a (modified) CTAB protocol or a rapid desalting method (MasterPure Complete DNA and RNA Purification Kit; Lucigen Corporation).

Total genomic DNA was isolated from 20 pooled adult *Daphnia* for each of the five EIC clonal lines using a CTAB extraction method (Doyle & Doyle, 1987).

Whole-genome amplification on resting eggs for resequencing

To isolate *Daphnia* resting eggs from the sediment each sediment layer was sieved using a sieve with 125 µm mesh size and small amounts of the remaining sediment were resuspended in distilled water. Ephippia were eye spotted under a stereomicroscope, counted, and transferred to 1.5 ml tubes. The water was removed and ephippia stored at −20 °C in the dark until further analysis.

The ephippia were then opened under a binocular with insect needles and tweezers previously treated under a clean bench (UV sterilization) and with DNase away (Thermo Fisher). Eggs that were already damaged, had an uneven shape or were orange, which is evidence for degradation, were discarded. The resting egg separated from the ephippial casing washed in 15 µl sterile 1× PBS and then transferred in 1 µl 1× PBS to a new tube with 2 µl fresh 1× PBS. The isolated eggs were stored at −80 °C at least overnight.

For WGA of single eggs, the REPLI-g Mini Kit (Qiagen) was used. This kit is enabling unbiased amplification of genomic loci via Multiple Displacement Amplification. The isolated resting eggs were thawed on ice and the whole genome was amplified following the manufacturer's protocol for amplification of genomic DNA from blood or cells. Briefly, denaturation buffer was added to the prepared resting eggs in 3 µl 1× PBS and amplified by REPLI-g Mini DNA Polymerase under isothermal conditions for 16 hours.

The amplified product was quantified on a Nanodrop spectrophotometer (Thermo Fisher) and with a Qubit Fluorometer (Thermo Fisher). Successful amplifications were purified with 0.4× Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove small fragments and eluted in 60 µl 1× TE buffer.

Fragments of the mitochondrial gene 16S rRNA gene were amplified to check successful amplification of *Daphnia* DNA using the universal cladoceran primers S1 and S2 (Schwenk et al., 1998) and a low presence of bacterial DNA using universal primers for the bacterial 16S rDNA gene (Nadkarni et al., 2002). Only samples with a successful amplification of the *Daphnia* 16S fragment and low amplification of the bacterial 16S fragment, indicating low bacterial contamination, were used for sequencing steps.

Library preparation and sequencing of resequencing samples

After quantification and quality control of the DNA using Nanodrop and Qubit instruments, libraries were prepared either directly in-house with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs), or at the sequencing company Novogene (Cambridge, UK). Resequencing (paired-end 150-bp reads) was then performed either at Novogene (UK) Company Limited or the Functional Genomics Center (ETH Zurich and University of Zurich) on Illumina NovaSeq 6000 and HiSeq4000 instruments. Details on the procedure used for each sample are provided in Table S3.2.

Genome assembly and annotation

We provide here a summarized version of the procedure used to assemble and annotate the genome. Details can be found in Supplementary Information on Chapter 3 – Supplementary materials and methods.

Raw data QC

Illumina reads were trimmed and the adapter removed using a combination of Trimmomatic 0.38 (Bolger et al., 2014), FastQC 0.11.7 (Andrews, 2010) and MultiQC 1.6 (Ewels et al., 2016) within autotrim 0.6.1 (Waldvogel et al., 2018). To filter out reads possibly originating from contamination from known sources (see below), a FastQ Screen like approach was chosen. In brief, the reads are separated by results of mapping behavior to different genomes. Positive controls consisted of genome data for other *Daphnia* species (dmagna-v2.4 and Daphnia_pulex_PA42_v3.0, see Supplementary Information on Chapter 3 – Supplementary materials and methods for accession numbers), and negative control, that is, sequences deemed undesirable for genome assembly consisted of genome data from human, bacteria, viruses, and the algae used to feed the batch cultures. The resulting database comprised 108,163 sequences (total sequence space 42.2 Gb). Both Illumina reads and PacBio subreads were mapped against the database with NextGenMap (Sedlazeck et al., 2013) and minimap2 (Li, 2018), respectively.

Reads did only pass the filtering if they either did not map to the database at all or had at least one hit against one of the two *Daphnia* genomes. Table M3.2 gives an overview of the effect of different filtering steps.

Assembly and contamination screening

All paired and unpaired contamination filtered Illumina reads as well as the contamination filtered PacBio reads were used as input for RA 0.2.1 (https://github.com/rvaser/ra, last accessed November 2021). Blobtools 1.0 (Laetsch & Blaxter, 2017) was used to screen the resulting assembly for possible unidentified contamination in the hybrid assembly. Briefly, bwa mem 0.7.17 (Li, 2013) was used to map Illumina reads back to the assembly and taxonomic assignment was done by sequence similarity search with BlastN 2.9.0+ (Camacho et al., 2009). Contamination with different bacteria was clearly identifiable, and contigs with coverage below 10× and/or GC content above 50% were removed. Additionally, PacBio reads mapping to these contigs were removed to minimize false scaffolding in further steps. The contig corresponding to the mitochondrial genome was identified after a BLAST search against available mitochondrial genomes for this species and removed from the assembly.

Scaffolding and gap closing

The blobtools filtered PacBio reads were used for scaffolding and gapclosing, which was conducted in three iterations. Each iteration consisted of a scaffolding step with SSPACE LongRead 1-1 (Boetzer & Pirovano, 2014), a gap closing step with LR Gapcloser (https://github.com/CAFS-bioinformatics/LR_Gapcloser; commit 156381a, last accessed November 2021), and a step to polish former gap parts with short reads using bwa mem 0.7.17-r1188 and

Pilon 1.23 (Walker et al., 2014) in a pipeline developed to this effect, wtdbg2-racon-pilon.pl 0.4 (https://github.com/schellt/wtdbg2-racon-pilon, last accessed November 2021).

Assembly quality assessment

Contiguity was analyzed with Quast 5.0.2 (Gurevich et al., 2013) at different stages of the assembly process. Further, mapping rate, coverage, and insert size distribution were assessed by mapping Illumina and PacBio reads with bwa mem and Minimap 2.17 respectively. To show absence of contamination in the assembly blobtools was run as above. The genome size was estimated by dividing the mapped nucleotides by the mode of the coverage distribution of the Illumina reads by backmap 0.3 (https://github.com/schellt/backmap, last accessed January 2021), resulting in 156.86 Mb (with the obtained assembly length amounting to 85% of this estimated length). Additionally, the genome size was estimated using a k-mer based approach by creating a histogram from raw Illumina reads with Jellyfish 1.1.12 (Marçais & Kingsford, 2011) and running the GenomeScope web application (http://qb.cshl.edu/genomescope/, last accessed November 2021) resulting in a genome size estimate of 150.6 Mb.

Completeness in terms of single copy core orthologs of the final scaffolds was assessed with BUSCO 3.0.2 (Simão et al., 2015), using the Arthropoda set (odb9).

Genome annotation

RepeatModeler 2.0 (Smit & Hubley, 2008–2015) was run to identify *D. galeata* specific repeats. The 1,115 obtained repeat families were combined with 237 *D. pulex* and one *D. pulicaria* repeat sequences from RepBase release 20181026 to create the final repeat library. The genome assembly was then soft masked with RepeatMasker 4.1.0 (Smit et al., 2013–2015), resulting in 21.9% of the assembly being masked.

Gene prediction models were produced with Augustus 3.3.2 (Stanke et al., 2008), GeneMark ET 4.48_3.60_lic (Lomsadze et al., 2005), and SNAP 2006-07-28 (Korf, 2004). The Augustus model was based on the soft masked assembly and the *D. galeata* transcriptome (HAFN01.1, Huylmans et al., 2016). The GeneMark model was obtained by first mapping trimmed RNAseq reads to the assembly with HISAT 2.1.0 (Kim et al., 2019) and then processing the resulting bam file with bam2hints and filterIntronsFindStrand.pl from Augustus to create a gff file with possible introns, which was finally fed into GeneMark.

The structural annotation was conducted in MAKER 2.31.10 (Holt & Yandell, 2011). Briefly, the unmasked genome assembly, the species' own transcriptome assembly as ESTs, the complete Swiss-Prot 2019_10 (UniProt Consortium, 2019) and the protein sequences resulting from *D. magna* (Lee et al., 2019), as well as *D. pulex* (Ye et al., 2017) genome annotations as protein evidence, were used as input for MAKER. In total, three iterations of MAKER with

retraining of the Augustus and SNAP model in between the iterations were conducted. The quality of the structural annotation was assessed by comparing values as number of genes, gene space, etc. to existing annotations for other *Daphnia* genomes. Furthermore, core orthologs from BUSCO's Arthropoda (odb9) set and conserved domain arrangements from the Arthropoda reference set of DOGMA 3.4 (Dohmen et al., 2016) were searched in the annotated protein set.

The functional annotation was conducted using InterProScan 5.39-77.0 (Jones et al., 2014) as well as a BLAST against the Swiss-Prot 2019_10.

Population samples

Raw data QC and contamination check

The quality of raw reads was checked using FastQC v0.11.5. Adapter trimming and quality filtering were performed using Trimmomatic v0.36 with the following parameters: ILLUMINACLIP: TruSeq3-PE.fa: 2:30:10 TRAILING: 20 SLIDINGWINDOW: 4:15 MINLEN: 70. For samples sequenced on a NovaSeq6000 instrument and presenting a typical polyG tail, the program fastp (Chen et al., 2018) was used for trimming as well. To assess contamination in the WGA samples FastQ Screen v0.14.0 with the bwa mapping option was used (Wingett & Andrews, 2018). A custom database was built to map trimmed reads against possible contaminants that included general common contaminants such as *Homo sapiens*, the UniVec reference database, a bacterial and a viral reference set as well as the *D. galeata* genome and *Acutodesmus obliquus* draft genome (see Supplementary Information on Chapter 3 – Supplementary materials and methods for accession numbers). Samples with <25% reads mapped to the *D. galeata* genome (and 25% contamination) were excluded from further analysis because the whole amplification of the resting egg most likely failed.

Mapping to reference genome and variant calling

The variant calling was performed within the Genome Analysis Toolkit (GATK v4.1.4.0; McKenna et al., 2010) program according to GATK4 best practices (Van der Auwera et al., 2013). The trimmed reads were mapped to the *D. galeata* genome using the BWA-MEM algorithm in BWA v0.7.17 with the -M parameter and adding read group identifiers for Picard compatibility (Li & Durbin, 2009). PCR duplicates were marked and filtered out in the BAM file using Picard v2.21.1 (http://broadinstitute.github.io/picard/, last accessed November 2021).

To call variants for each sample, GATK HaplotypeCaller in GATK was used with the –emitRefConfidence GVCF option resulting in a genomic variant call format (gVCF) file with information on each position for each individual (Poplin et al., 2018). All gVCF files were consolidated using CombineGVCFs and joint genotyping was performed with

GenotypeGVCFs. The VCF file was filtered to include only SNPs and hard filtering was performed to remove variants with a QualByDepth <10, StrandOddsRatio >3, FisherStrand >60, mapping quality <40, MappingQualityRankSumTest <−8, and ReadPosRankSumTest <−5.

Subsequently, we removed sites with either very high coverage (>450) or for which genotypes were missing for more than 20% of the individuals using VCFtools v0.1.5 (Danecek et al., 2011). The final SNP data set for downstream analyses included 3,240,339 SNPs across the 49 samples.

In addition, GenotypeGVCFs was run with the –include-non-variant-sites option to output all variant as well as invariant genotyped sites. The final invariant data set included 127,530,229 sites after removing indels and multiallelic sites with BCFtools v1.9 (Li, 2011) and is used for population genomic analysis to be able to calculate the total number of genotyped sites (variant and invariant) within a genomic window.

As we mapped all different species to the reference *D. galeata* genome, we assessed possible reference bias by checking the distribution of reference and alternative alleles observed at heterozygous genotypes based on Pinsky et al. (2021). We pooled all genotypes which were unambiguously assigned to either of the parental species clusters *D. galeata*, *D. cucullata*, and *D. longispina* as was done for the population genomic parameters (Table S3.2) or classified as hybrids using ADMIXTURE inference. Without reference bias, we would expect that in heterozygous genotypes the reference and the alternative allele are on an average represented by 50% of the reads. An indication of reference bias would be that the *D. galeata* reference allele would be more frequent.

Phylogenetic and population genetics inferences

Mitochondrial genome assemblies and phylogenetic analyses

All reads were used to produce mitochondrial genome assemblies using the “de novo assembly” and “find mitochondrial scaffold” modules provided in MitoZ v2.4 with default settings (Meng et al., 2019). For some samples, this was not sufficient and we used two approaches to recover a complete mitogenome: either the mitochondrial baiting and iterative mapping implemented in MITObim v1.9.1 (Hahn et al., 2013) with the *D. galeata* mitochondrial reference genome or the modified baiting and iterative mapping in GetOrganelle v1.7.1 (Jin et al., 2020) with the animal database and k-mer values set to 21, 45, 65, 85, and 105. The procedure used for each data set is given in Table S3.5.

We annotated the mitochondrial genome assemblies with the mitochondrial annotation web server MITOS2 (Bernt et al., 2013) using the mitochondrial codon code 05 for invertebrates.

Automated genome annotation identified 13 PCGs, two rRNA genes, and 22 transfer RNA genes (tRNAs). Initially, the mitochondrial genes (PCGs and rRNAs, Table S3.6) were individually aligned with MUSCLE v3.8.1551 (Edgar, 2004) and visually checked for their quality. The mitochondrial genome assemblies with discrepancies, that is, a lot of missing data and/or split features were excluded from further analysis. The final data set (Table S3.6) included 44 mitochondrial genomes from this study and the previously published mitochondrial genome of *Daphnia laevis* (Martins Ribeiro et al., 2019, accession number: NC_045243.1). The mitochondrial genes of the final data set were individually realigned with MUSCLE v3.8.1551 (Edgar, 2004) and MACSE v2.05 (Ranwez et al., 2018) and concatenated into a mitochondrial DNA matrix (Table S3.6) using SequenceMatrix v1.8.1 (Vaidya et al., 2011). During this step, we used MACSE v2.05 to realign PCG genes keeping the information about codon position (gene partitioning) and to remove STOP codons. The final data set consisted of the concatenation matrix of the 13 PCGs and the two structural rRNA genes. With this alignment, phylogenetic trees were reconstructed using IQ-TREE v1.6.12 (Nguyen et al., 2015). We initially partitioned the alignment into a full partition model, that is, each gene and all three codon positions for PCGs, and then ran IQ-TREE with partition analyses (-spp, Chernomor et al., 2016), ModelFinder (-m MFP+MERGE, Kalyaanamoorthy et al., 2017) and 10,000 ultrafast bootstrap (-bb 10,000, Hoang et al., 2018) and SH-like approximate likelihood ratio test (-alrt 10,000, Guindon et al., 2010) replicates. The resulting trees were visualized in R (R Core Team, 2017) using the multifunctional phylogenetics package phytools (Revell, 2012).

Ancestry and population structure

A principal component analysis was conducted in R v3.6.2 (R Core Team, 2017) with the package SNPRelate v1.20.1 (Zheng et al., 2012). Linkage disequilibrium (LD) was calculated within a 500-kb sliding window and LD-pruned for r^2 values >0.5 before conducting the PCA for all sites using the snpgdsPCA function with default settings. The relatively large LD value was chosen because clonal reproduction and the overlap of generations due to diapause leads to increased LD in *Daphnia* (Brede et al., 2009).

Genetic admixture was estimated using ADMIXTURE v1.3.0 (Alexander & Lange, 2011). The SNP set VCF file was converted to BED format using plink v1.90b6.13 (Chang et al., 2015). The log-likelihood values were estimated for one to five genetic clusters (*K*) of ancestral populations and admixture analysis was run for the most appropriate *K* value with 10-fold cross-validation. We also conducted the PCA and Admixture analysis using PCAngsd implemented in ANGSD and NgsAdmix, respectively (Korneliussen et al., 2014) to take genotype likelihoods into account (details in Supplementary Information on Chapter 3 –

Supplementary materials and methods). The results did not differ substantially and are shown in Figure S3.3.

However, using such a population genetic clustering approach to estimate ancestry coefficients is not directly equivalent to the proportion of hybrid ancestry in each individual and should be interpreted with caution (Kong & Kubatko, 2021; Lawson et al., 2018). The results of the ADMIXTURE analysis suggested that the data set included hybrids between *D. longispina* and *D. galeata* as well as *D. cucullata* and *D. galeata*. We then followed the "ancestry painting" procedure outlined in Barth et al. (2020) and Runemark et al. (2018), and classified sites according to their F_{ST} values when comparing parental species sets. Unlike the PCA and the admixture analysis, this approach requires the user to define parental genotypes; the individuals belonging to these sets are indicated with stars in Figure 3.1C. Fixed sites are those where a specific allele is fixed in all individuals belonging to one parental species and another allele fixed in the other parental species. To show the ancestry of the hybrid individuals each fixed site was plotted in an "ancestry painting" if at least 80% of genotypes were complete using available ruby scripts (https://github.com/mmatschiner/tutorials/tree/master/analysis_of_introgression_with_snp_data, last accessed November 2021). These scripts calculate the heterozygosity of each individual and visualize regions that are possibly affected by introgression. The mitochondrial genome assembly from each individual was used to determine the maternal species and the proportion of the genome derived from the maternal species was then calculated for each hybrid. For gal×cuc hybrids the hybrid index scale ranges from 0 (gal) to 1 (cuc) and for gal×long hybrids from 0 (long) to 1 (gal).

Window-based population parameters

To assess genome-wide genetic differentiation between the clusters identified with admixture, we calculated nucleotide diversity (π), between-taxon differentiation (F_{ST}), and between-taxon divergence (d_{xy}) using the Python script popgenWindows.py (github.com/simonhmartin/genomics_general release 0.3, Martin et al., 2020, last accessed November 2021) with a sliding 100-kb window, a step size of 10 kb and at least 20-kb genotyped sites within each window. To compare species pairs, we only considered individuals assigned to parental species based on ADMIXTURE results (Table S3.2). In addition, we also calculated these parameters using one randomly chosen individual from each population per species to check if the estimates are biased because of the overrepresentation of some populations in a species group (Table S3.2).

Sets of outlier windows were defined as those with F_{ST} values in the upper 95th percentile of the distribution for each of the three pairwise comparisons. Further, the genes in these windows were extracted using the annotation file. We used a randomization approach to

assess whether the observed intersections (i.e., outlier F_{ST} windows occurring in both species) between all seven possible species comparisons are larger or smaller than expected by chance. For this, we randomly drew the observed number of windows, respectively genes from the total number of 10-kb windows in the assembly (13,330), respectively the total number of annotated genes (15,845) without replacement and calculated the intersections for all possible comparisons. We compared the resulting intersections from 1,000 replicates with the observed values (Figure S3.5).

Inferring introgression

To identify admixture among three populations, we calculated the f_3 statistic with ADMIXTOOLS v 7.0 (Patterson et al., 2012) implemented in the admixr package in R (Petr et al., 2019). We used two parental source populations (A and B) and the target population (C) in the form (C; A, B). Significantly negative f_3 statistic indicates that population C is a mixture of populations A and B or closely related populations.

Local ancestry inference

To prepare the SNP set, Beagle v4.1 was used to phase and impute genotypes with 10,000 bp step size and 1,000-bp overlapping sliding windows (Browning & Browning, 2009). Local ancestry inference was conducted with Loter (Dias-Alves et al., 2018) which infers the origin of each SNP in an admixed individual from two ancestral source populations and does not require additional biological parameters. The respective two parental species populations were used to reconstruct the ancestry tracts of the three putative *galeata*×*cucullata* hybrid individuals and 11 putative *galeata*×*longispina* hybrid individuals using Loter with default settings.

Data accessibility

Genome assembly, annotation, and read data (Illumina and PacBio) for the genotype M5 are stored under accession number PRJEB42807. Short-read data from resequencing are available in the European Nucleotide Archive under accession numbers ERS5080327–ERS5080375, ERS4993274, and ERS4993282. The annotation and genome assembly used in the present analysis are deposited in Zenodo (doi:10.5281/zenodo.4479324), together with supplementary information on the mitochondrial tree (alignment file).

Acknowledgments

We thank LOEWE-TBG for providing sequencing funds. T.H. and M.M. were supported by the grant “SeeWandel: Life in Lake Constance—the past, present and future” within the framework of the Interreg V programme “Alpenrhein-Bodensee-Hochrhein (Germany/Austria/Switzerland/ Liechtenstein),” which funds are provided by the European Regional Development Fund as

well as the Swiss Confederation and cantons. M.M. was supported by the Austrian Science Fund (FWF): P29667-B25 and J 3774. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck. Some of the data produced and analyzed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. We thank Jae-Seong Lee and Zhiqiang Ye for giving us access to genome annotations for *D. magna* (SK strain) and *D. pulex* (PA42 strain), respectively. Miklós Bálint provided sediment samples and isotope data for the Arendsee lake. We thank Michael Matschiner for his help with ancestry painting and two anonymous reviewers for their comments on a previous version of the manuscript.

Chapter 3 References

Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Adamowicz, S. J., Petrusek, A., Colbourne, J. K., Hebert, P. D. N., & Witt, J. D. S. (2009). The scale of divergence: A phylogenetic appraisal of intercontinental allopatric speciation in a passively dispersed freshwater zooplankton genus. *Molecular Phylogenetics and Evolution*, *50*(3), 423–436. https://doi.org/10.1016/j.ympev.2008.11.026

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Alric, B., Möst, M., Domaizon, I., Pignol, C., Spaak, P., & Perga, M.-E. (2016). Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology*, *29*(4), 720–735. https://doi.org/10.1111/jeb.12820

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnold, M. L., & Martin, N. H. (2009). Adaptation by introgression. *Journal of Biology*, *8*, 82. https://doi.org/10.1186/jbiol176

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Barth, J. M. I., Gubili, C., Matschiner, M., Tørresen, O. K., Watanabe, S., Egger, B., Han, Y.-S., Feunteun, E., Sommaruga, R., Jehle, R., & Schabetsberger, R. (2020). Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15099-x

Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Beaton, M. J., & Hebert, P. D. N. (1994). Variation in chromosome numbers of *Daphnia* (Crustacea, Cladocera). *Hereditas*, *120*(3), 275–279. https://doi.org/10.1111/j.1601-5223.1994.00275.x

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Billiones, R., Brehm, M., Klee, J., & Schwenk, K. (2004). Genetic identification of *Hyalodaphnia* species and interspecific hybrids. *Hydrobiologia*, *526*(1), 43–53. https://doi.org/10.1023/B:HYDR.0000041615.65087.06

Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, *15*(1), 211. https://doi.org/10.1186/1471-2105-15-211

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences*, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Brede, N., Thielsch, A., Sandrock, C., Spaak, P., Keller, B., Streit, B., & Schwenk, K. (2006). Microsatellite markers for European *Daphnia*. *Molecular Ecology Notes*, *6*(2), 536–539. https://doi.org/10.1111/j.1471-8286.2005.01218.x

Browning, B. L., & Browning, S. R. (2009). A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *American Journal of Human Genetics*, *84*(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sætre, G.-P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, *25*(11), 1656–1665. https://doi.org/10.1101/gr.196485.115

Butlin, R. K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., Coyne, J. A., Galindo, J., Grahame, J. W., Hollander, J., Kemppainen, P., Martínez-Fernández, M., Panova, M., Quesada, H., Johannesson, K., & Rolán-Alvarez, E. (2014). Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution*, *68*(4), 935–949. https://doi.org/10.1111/evo.12329

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Canestrelli, D., Bisconti, R., Chiocchio, A., Maiorano, L., Zampiglia, M., & Nascetti, G. (2017). Climate change promotes hybridisation between deeply divergent species. *PeerJ*, *5*, e3072. https://doi.org/10.7717/peerj.3072

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chernomor, O., Von Haeseler, A., & Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, *65*(6), 997–1008. https://doi.org/10.1093/sysbio/syw037

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cáceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., … Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, *331*(6017), 555–561. https://doi.org/10.1126/science.1197761

Cordellier, M., Wojewodzic, M. W., Wessels, M., Kuster, C., & Von Elert, E. (2021). Next-generation sequencing of DNA from resting eggs: Signatures of eutrophication in a lake's sediment. *Zoology*, *145*, 125895. https://doi.org/10.1016/j.zool.2021.125895

Cornetti, L., Fields, P. D., Van Damme, K., & Ebert, D. (2019). A fossil-calibrated phylogenomic analysis of *Daphnia* and the Daphniidae. *Molecular Phylogenetics and Evolution*, *137*, 250–262. https://doi.org/10.1016/j.ympev.2019.05.018

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D., & Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proceedings of the National Academy of Sciences*, *98*(11), 6256–6260. https://doi.org/10.1073/pnas.111606798

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dias-Alves, T., Mairal, J., & Blum, M. G. B. (2018). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. *Molecular Biology and Evolution*, *35*(9), 2318–2326. https://doi.org/10.1093/molbev/msy126

Dlouhá, Š., Thielsch, A., Kraus, R. H. S., Seda, J., Schwenk, K., & Petrusek, A. (2010). Identifying hybridizing taxa within the *Daphnia longispina* species complex: A comparison of genetic methods and phenotypic approaches. *Hydrobiologia*, *643*(1), 107–122. https://doi.org/10.1007/s10750-010-0128-8

Doellman, M. M., Ragland, G. J., Hood, G. R., Meyers, P. J., Egan, S. P., Powell, T. H. Q., Lazorchak, P., Glover, M. M., Tait, C., Schuler, H., Hahn, D. A., Berlocher, S. H., Smith, J. J., Nosil, P., & Feder, J. L. (2018). Genomic Differentiation during Speciation-with-Gene-Flow: Comparing Geographic and Host-Related Variation in Divergent Life History Adaptation in *Rhagoletis pomonella*. *Genes*, *9*(5), 262. https://doi.org/10.3390/genes9050262

Dohmen, E., Kremer, L. P. M., Bornberg-Bauer, E., & Kemena, C. (2016). DOGMA: Domain-based transcriptome and proteome quality assessment. *Bioinformatics*, *32*(17), 2577–2581. https://doi.org/10.1093/bioinformatics/btw231

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Dziuba, M. K., Herdegen-Radwan, M., Pluta, E., Wejnerowski, Ł., Szczuciński, W., & Cerbin, S. (2020). Temperature increase altered *Daphnia* community structure in artificially heated lakes: A potential scenario for a warmer future. *Scientific Reports*, *10*(1), 13956. https://doi.org/10.1038/s41598-020-70294-6

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Figuerola, J., Green, A. J., & Michot, T. C. (2005). Invertebrate eggs can fly: Evidence of waterfowl-mediated gene flow in aquatic invertebrates. *The American Naturalist*, *165*(2), 274–280. https://doi.org/10.1086/427092

Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*(16), 4074–4088. https://doi.org/10.1111/mec.12750

Folk, R. A., Mandel, J. R., & Freudenstein, J. V. (2017). Ancestral Gene Flow and Parallel Organellar Genome Capture Result in Extreme Phylogenomic Discord in a Lineage of Angiosperms. *Systematic Biology*, *66*(3), 320–337. https://doi.org/10.1093/sysbio/syw083

Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene-flow in a mosaic hybrid zone: Is local introgression adaptive? *Genetics*, *197*(3), 939–951. https://doi.org/10.1534/genetics.114.161380

Frisch, D., Green, A. J., & Figuerola, J. (2007). High dispersal capacity of a broad spectrum of aquatic invertebrates via waterbirds. *Aquatic Sciences*, *69*(4), 568–574. https://doi.org/10.1007/s00027-007-0915-0

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368. https://doi.org/10.1111/ele.12237

Gannon, J. E., & Stemberger, R. S. (1978). Zooplankton (Especially Crustaceans and Rotifers) as Indicators of Water Quality. *Transactions of the American Microscopical Society*, *97*(1), 16–35. https://doi.org/10.2307/3225681

Gliwicz, Z. M. (1990). Why do cladocerans fail to control algal blooms? In R. D. Gulati, E. H. R. R. Lammens, M.-L. Meijer, & E. Van Donk (Eds.), *Biomanipulation Tool for Water Management* (pp. 83–97). Springer Netherlands. https://doi.org/10.1007/978-94-017-0924-8_8

Griebel, J., Gießler, S., Yin, M., & Wolinska, J. (2016). Parental and hybrid *Daphnia* from the *D. longispina* complex: Long-term dynamics in genetic structure and significance of overwintering modes. *Journal of Evolutionary Biology*, *29*(4), 810–823. https://doi.org/10.1111/jeb.12828

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129. https://doi.org/10.1093/nar/gkt371

Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, *56*(3), 337–349. https://doi.org/10.1038/hdy.1986.55

Herrmann, M., Henning-Lucass, N., Cordellier, M., & Schwenk, K. (2017). A genotype–phenotype association approach to reveal thermal adaptation in *Daphnia galeata*. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, *327*(1), 53–65. https://doi.org/10.1002/jez.2070

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. https://doi.org/10.1186/1471-2105-12-491

Huang, D. I., Hefer, C. A., Kolosova, N., Douglas, C. J., & Cronk, Q. C. B. (2014). Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, *Populus balsamifera* and *P. trichocarpa* (Salicaceae). *The New Phytologist*, *204*(3), 693–703. https://doi.org/10.1111/nph.12956

Huylmans, A. K., López Ezquerra, A., Parsch, J., & Cordellier, M. (2016). De Novo Transcriptome Assembly and Sex-Biased Gene Expression in the Cyclical Parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution*, *8*(10), 3120–3139. https://doi.org/10.1093/gbe/evw221

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., DePamphilis, C. W., Yi, T.-S., & Li, D.-Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*(1), 241. https://doi.org/10.1186/s13059-020-02154-5

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kaiser, T. S., Von Haeseler, A., Tessmar-Raible, K., & Heckel, D. G. (2021). Timing strains of the marine insect *Clunio marinus* diverged and persist with gene flow. *Molecular Ecology*, *30*(5), 1264–1280. https://doi.org/10.1111/mec.15791

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Keller, B., Wolinska, J., Manca, M., & Spaak, P. (2008). Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2943–2952. https://doi.org/10.1098/rstb.2008.0044

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Klüttgen, B., Dülmer, U., Engels, M., & Ratte, H. T. (1994). ADaM, an artificial freshwater for the culture of zooplankton. *Water Research*, *28*(3), 743–746. https://doi.org/10.1016/0043-1354(94)90157-0

Kong, S., & Kubatko, L. S. (2021). Comparative Performance of Popular Methods for Hybrid Detection using Genomic Data. *Systematic Biology*, *70*(5), 891–907. https://doi.org/10.1093/sysbio/syaa092

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*(1), 59. https://doi.org/10.1186/1471-2105-5-59

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2018). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, *18*(1), 118–127. https://doi.org/10.1111/1755-0998.12720

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Lampert, W., & Sommer, U. (2007). *Limnoecology: The Ecology of Lakes and Streams* (Second Edition). Oxford University Press.

Lawson, D. J., Van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, *9*(1), 3258. https://doi.org/10.1038/s41467-018-05257-7

Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, *210*, 69–84. https://doi.org/10.1016/j.aquatox.2019.02.009

Lee-Yaw, J. A., Grassa, C. J., Joly, S., Andrew, R. L., & Rieseberg, L. H. (2019). An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytologist*, *221*(1), 515–526. https://doi.org/10.1111/nph.15386

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Limburg, P. A., & Weider, L. J. (2002). 'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*(1488), 281–287. https://doi.org/10.1098/rspb.2001.1868

Llopart, A., Herrig, D., Brud, E., & Stecklein, Z. (2014). Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Molecular Ecology*, *23*(5), 1124–1136. https://doi.org/10.1111/mec.12678

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506. https://doi.org/10.1093/nar/gki937

Ma, X., Hu, W., Smilauer, P., Yin, M., & Wolinska, J. (2019). *Daphnia galeata* and *D. dentifera* are geographically and ecologically separated whereas their hybrids occur in intermediate habitats: A survey of 44 Chinese lakes. *Molecular Ecology*, *28*(4), 785–802. https://doi.org/10.1111/mec.14991

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Marková, S., Dufresne, F., Manca, M., & Kotlík, P. (2013). Mitochondrial Capture Misleads about Ecological Speciation in the *Daphnia pulex* Complex. *PLoS One*, *8*(7), e69497. https://doi.org/10.1371/journal.pone.0069497

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, *17*(2), e2006288. https://doi.org/10.1371/journal.pbio.2006288

Martin, S. H., Singh, K. S., Gordon, I. J., Omufwoko, K. S., Collins, S., Warren, I. A., Munby, H., Brattström, O., Traut, W., Martins, D. J., Smith, D. A. S., Jiggins, C. D., Bass, C., & Ffrench-Constant, R. H. (2020). Whole-chromosome hitchhiking driven by a male-killing endosymbiont. *PLoS Biology*, *18*(2), e3000610. https://doi.org/10.1371/journal.pbio.3000610

Martins Ribeiro, M., Facchin, S., Pereira, A. H., Kalapothakis, E., Xu, S., Han, B.-P., Dumont, H. J., & Rietzler, A. C. (2019). Mitogenome of *Daphnia laevis* (Cladocera, Daphniidae) from Brazil. *Mitochondrial DNA Part B*, *4*(1), 194–196. https://doi.org/10.1080/23802359.2018.1545547

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Melo-Ferreira, J., Vilela, J., Fonseca, M. M., Da Fonseca, R. R., Boursot, P., & Alves, P. C. (2014). The Elusive Nature of Adaptive Mitochondrial DNA Evolution of an Arctic Lineage Prone to Frequent Introgression. *Genome Biology and Evolution*, *6*(4), 886–896. https://doi.org/10.1093/gbe/evu059

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63–e63. https://doi.org/10.1093/nar/gkz173

Miner, B. E., De Meester, L., Pfrender, M. E., Lampert, W., & Hairston, N. G. (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1735), 1873–1882. https://doi.org/10.1098/rspb.2011.2404

Möst, M. (2013). *Environmental change and its impact on hybridising Daphnia species complexes* [Ph.D. thesis, ETH Zürich]. https://doi.org/10.3929/ethz-a-010076219

Nadkarni, M. A., Martin, F. E., Jacques, N. A., & Hunter, N. (2002). Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology*, *148*(1), 257–266. https://doi.org/10.1099/00221287-148-1-257

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J. K., Pfrender, M. E., & Weider, L. J. (2013). The evolutionary time machine: Using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution*, *28*(5), 274–282. https://doi.org/10.1016/j.tree.2013.01.009

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Petr, M., Vernot, B., & Kelso, J. (2019). admixr—R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*, *35*(17), 3194–3195. https://doi.org/10.1093/bioinformatics/btz030

Petrusek, A., Hobæk, A., Nilssen, J. P., Skage, M., Černý, M., Brede, N., & Schwenk, K. (2008). A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta*, *37*(5), 507–519. https://doi.org/10.1111/j.1463-6409.2008.00336.x

Petrusek, A., Seda, J., Macháček, J., Ruthová, Š., & Šmilauer, P. (2008). *Daphnia* hybridization along ecological gradients in pelagic environments: The potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2931–2941. https://doi.org/10.1098/rstb.2008.0026

Petrusek, A., Thielsch, A., & Schwenk, K. (2012). Mitochondrial sequence variation suggests extensive cryptic diversity within the Western Palearctic *Daphnia longispina* complex. *Limnology and Oceanography*, *57*(6), 1838–1845. https://doi.org/10.4319/lo.2012.57.6.1838

Pietrzak, B., & Ślusarczyk, M. (2006). The fate of the ephippia – *Daphnia* dispersal in time and space. *Polish Journal of Ecology*, *54*(4), 709–714.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., Baalsrud, H. T., Brieuc, M. S. O., Kjesbu, O. S., Godiksen, J. A., Barth, J. M. I., Matschiner, M., Stenseth, N. C., Jakobsen, K. S., Jentoft, S., & Star, B. (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences*, *118*(15), e2025453118. https://doi.org/10.1073/pnas.2025453118

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. https://doi.org/10.1101/201178

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

Rafajlović, M., Emanuelsson, A., Johannesson, K., Butlin, R. K., & Mehlig, B. (2016). A universal mechanism generating clusters of differentiated loci during divergence-with-migration. *Evolution*, *70*(7), 1609–1621. https://doi.org/10.1111/evo.12957

Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N., & Delsuc, F. (2018). MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution*, *35*(10), 2582–2584. https://doi.org/10.1093/molbev/msy159

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494. https://doi.org/10.1038/nature08365

Rellstab, C., Keller, B., Girardclos, S., Anselmetti, F. S., & Spaak, P. (2011). Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. *Limnology and Oceanography*, *56*(1), 292–302. https://doi.org/10.4319/lo.2011.56.1.0292

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S. R., De Carvalho, C. F., Safran, R. J., Sandoval, C. P., Feder,

J., Gries, R., Crespi, B. J., Gries, G., Gompert, Z., & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, *1*, 0082. https://doi.org/10.1038/s41559-017-0082

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., Elgvin, T. O., & Sætre, G.-P. (2018). Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution*, *2*(3), 549–556. https://doi.org/10.1038/s41559-017-0437-7

Rusek, J., Ayan, G. B., Turko, P., Tellenbach, C., Giessler, S., Spaak, P., & Wolinska, J. (2015). New possibilities arise for studies of hybridization: SNP-based markers for the multi-species *Daphnia longispina* complex derived from transcriptome data. *Journal of Plankton Research*, *37*(3), 626–635. https://doi.org/10.1093/plankt/fbv028

Sarver, B. A. J., Herrera, N. D., Sneddon, D., Hunter, S. S., Settles, M. L., Kronenberg, Z., Demboski, J. R., Good, J. M., & Sullivan, J. (2021). Diversification, Introgression, and Rampant Cytonuclear Discordance in Rocky Mountains Chipmunks (Sciuridae: *Tamias*). *Systematic Biology*, *70*(5), 908–921. https://doi.org/10.1093/sysbio/syaa085

Schreiber, D., & Pfenninger, M. (2021). Genomic divergence landscape in recurrently hybridizing *Chironomus* sister taxa suggests stable steady state between mutual gene flow and isolation. *Evolution Letters*, *5*(1), 86–100. https://doi.org/10.1002/evl3.204

Schwenk, K. (1993). Interspecific hybridization in *Daphnia*: Distinction and origin of hybrid matrilines. *Molecular Biology and Evolution*, *10*(6), 1289–1302. https://doi.org/10.1093/oxfordjournals.molbev.a040076

Schwenk, K., Bijl, M., & Menken, S. B. J. (2001). Experimental interspecific hybridization in *Daphnia*. *Hydrobiologia*, *442*(1), 67–73. https://doi.org/10.1023/A:1017594325506

Schwenk, K., Posada, D., & Hebert, P. D. N. (2000). Molecular systematics of European *Hyalodaphnia*: The role of contemporary hybridization in ancient species. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1455), 1833–1842.

Schwenk, K., Sand, A., Boersma, M., Brehm, M., Mader, E., Offerhaus, D., & Spaak, P. (1998). Genetic markers, genealogies and biogeographic patterns in the cladocera. *Aquatic Ecology*, *32*(1), 37–51. https://doi.org/10.1023/A:1009939901198

Sedlazeck, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, *29*(21), 2790–2791. https://doi.org/10.1093/bioinformatics/btt468

Seidendorf, B., Boersma, M., & Schwenk, K. (2007). Evolutionary stoichiometry: The role of food quality for clonal differentiation and hybrid maintenance in a *Daphnia* species complex. *Limnology and Oceanography*, *52*(1), 385–394. https://doi.org/10.4319/lo.2007.52.1.0385

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Skage, M., Hobæk, A., Ruthová, Š., Keller, B., Petrusek, A., Seda, J., & Spaak, P. (2007). Intra-specific rDNA-ITS restriction site variation and an improved protocol to distinguish species and hybrids in the *Daphnia longispina* complex. *Hydrobiologia*, *594*(1), 19–32. https://doi.org/10.1007/s10750-007-9090-5

Slager, D. L., Epperly, K. L., Ha, R. R., Rohwer, S., Wood, C., Hemert, C. V., & Klicka, J. (2020). Cryptic and extensive hybridization between ancient lineages of American crows. *Molecular Ecology*, *29*(5), 956–969. https://doi.org/10.1111/mec.15377

Sloan, D. B., Havird, J. C., & Sharbrough, J. (2017). The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*(8), 2212–2236. https://doi.org/10.1111/mec.13959

Smit, A. F. A., & Hubley, R. (2008–2015). *RepeatModeler Open-1.0.* Accessed November 2021. https://www.repeatmasker.org/RepeatModeler/

Smit, A. F. A., Hubley, R., & Green, P. (2013–2015). *RepeatMasker Open-4.0.* Accessed November 2021. https://www.repeatmasker.org/RepeatMasker/

Spaak, P., Denk, A., Boersma, M., & Weider, L. J. (2004). Spatial and temporal patterns of sexual reproduction in a hybrid *Daphnia* species complex. *Journal of Plankton Research*, *26*(6), 625–635. https://doi.org/10.1093/plankt/fbh064

Spaak, P., Fox, J., & Hairston, N. G. (2012). Modes and mechanisms of a *Daphnia* invasion. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2936–2944. https://doi.org/10.1098/rspb.2012.0280

Spaak, P., & Hoekstra, J. R. (1997). Fish predation on a *Daphnia* hybrid species complex: A factor explaining species coexistence? *Limnology and Oceanography*, *42*(4), 753–762. https://doi.org/10.4319/lo.1997.42.4.0753

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stephens, J. D., Rogers, W. L., Mason, C. M., Donovan, L. A., & Malmberg, R. L. (2015). Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. *American Journal of Botany*, *102*(6), 910–920. https://doi.org/10.3732/ajb.1500031

Taylor, D. J., Hebert, P. D. N., & Colbourne, J. K. (1996). Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Molecular Phylogenetics and Evolution*, *5*(3), 495–510. https://doi.org/10.1006/mpev.1996.0045

Thielsch, A., Brede, N., Petrusek, A., De Meester, L., & Schwenk, K. (2009). Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Molecular Ecology*, *18*(8), 1616–1628. https://doi.org/10.1111/j.1365-294X.2009.04130.x

Thielsch, A., Knell, A., Mohammadyari, A., Petrusek, A., & Schwenk, K. (2017). Divergent clades or cryptic species? Mito-nuclear discordance in a *Daphnia* species complex. *BMC Evolutionary Biology*, *17*, 227. https://doi.org/10.1186/s12862-017-1070-4

Thielsch, A., Völker, E., Kraus, R. H. S., & Schwenk, K. (2012). Discrimination of hybrid classes using cross-species amplification of microsatellite loci: Methodological challenges and solutions in *Daphnia*. *Molecular Ecology Resources*, *12*(4), 697–705. https://doi.org/10.1111/j.1755-0998.2012.03142.x

Tollrian, R., & Harvell, C. D. (1999). *The Ecology and Evolution of Inducible Defenses*. Princeton University Press.

Trentini, M. (1980). Chromosome numbers of nine species of Daphniidae (Crustacea, Cladocera). *Genetica*, *54*(2), 221–223. https://doi.org/10.1007/BF00055995

UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vergilino, R., Markova, S., Ventura, M., Manca, M., & Dufresne, F. (2011). Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. *Molecular Ecology*, *20*(6), 1191–1207. https://doi.org/10.1111/j.1365-294X.2011.05004.x

Waldvogel, A.-M., Wieser, A., Schell, T., Patel, S., Schmidt, H., Hankeln, T., Feldmeyer, B., & Pfenninger, M. (2018). The genomic footprint of climate adaptation in *Chironomus riparius*. *Molecular Ecology*, *27*(6), 1439–1456. https://doi.org/10.1111/mec.14543

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, *7*, 1338. https://doi.org/10.12688/f1000research.15931.1

Wolf, H. G. (1987). Interspecific hybridization between *Daphnia hyalina*, *D. galeata*, and *D. cucullata* and seasonal abundances of these species and their hybrids. *Hydrobiologia*, *145*(1), 213–217. https://doi.org/10.1007/BF02530282

Wolf, H. G., & Mort, M. A. (1986). Inter-specific hybridization underlies phenotypic variability in *Daphnia* populations. *Oecologia*, *68*(4), 507–511. https://doi.org/10.1007/BF00378763

Ye, Z., Xu, S., Spitze, K., Asselman, J., Jiang, X., Ackerman, M. S., Lopez, J., Harker, B., Raborn, R. T., Thomas, W. K., Ramsdell, J., Pfrender, M. E., & Lynch, M. (2017). A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes, Genomes, Genetics*, *7*(5), 1405–1416. https://doi.org/10.1534/g3.116.038638

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Yin, M., Gießler, S., Griebel, J., & Wolinska, J. (2014). Hybridizing *Daphnia* communities from ten neighbouring lakes: Spatio-temporal dynamics, local processes, gene flow and invasiveness. *BMC Evolutionary Biology*, *14*(1), 80. https://doi.org/10.1186/1471-2148-14-80

Zehnder, A., & Gorham, P. R. (1960). Factors influencing the growth of *Microcystis aeruginosa* Kütz. Emend. Elenkin. *Canadian Journal of Microbiology*, *6*(6), 645–660. https://doi.org/10.1139/m60-077

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Chapter 4 – Patterns of recurrent hybridization and genomic architecture through time in the *Daphnia longispina* species complex

Jana Nickel[1], Thijs Janzen[2], Tilman Schell[3], Markus Möst[4], Mathilde Cordellier[1,5]

[1] Institute of Cell and Systems Biology of Animals, University of Hamburg, Germany

[2] Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

[3] LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG), Frankfurt am Main, Germany

[4] Research Department for Limnology, University of Innsbruck, Mondsee, Austria

[5] Revvity Cellular Technologies, Hamburg, Germany

This chapter is in preparation for submission to *Molecular Ecology*.

Abstract

Closely related species that have recently come into contact due to ecological or anthropogenic changes can form hybrid zones. The genomes of hybrids in such zones exhibit varying levels of ancestry from their parental species that are determined and shaped by recombination and selection. Within the *Daphnia longispina* species complex, a keystone zooplankton group in freshwater ecosystems, three species occupy ecologically distinct niches and have been shown to hybridize extensively. These species produce resting eggs that accumulate in lake sediments over time, thus providing a direct record of past hybridization events. Here, we analyzed 68 whole genomes from single resting eggs collected from different time periods as well as individuals collected from the water column. We provide an improved chromosome-scale *Daphnia galeata* genome assembly to investigate genome-wide ancestry variation and introgression in two populations with different ecological conditions. We applied a novel junctions approach, which leverages how recombination breaks up genomic blocks of contiguous ancestry over time, to this data set to estimate the timing of hybridization events and the distribution of ancestry across the genome. We uncovered recurrent hybridization and introgression in different populations and between different species and showed high variation in hybrid ancestry among individuals and across the genome. Our work reveals the complexity of hybridization and introgression over several decades with unprecedented resolution, and the unique outcome of hybridization in a freshwater mosaic hybrid zone not previously described in a natural hybrid population.

Keywords: hybrid zone, introgression, population genetic time-series, resting eggs

Introduction

In hybrid zones, genomes are formed by admixture of closely related parental species and are shaped by neutral processes, recombination, or selection (Barton & Gale, 1993). Using population genomics approaches across a wide range of taxa allows us to better understand the causes and frequency of hybridization, how the ancestry composition of hybrids varies within populations, and the short- and long-term evolutionary consequences of hybridization (Gompert et al., 2017; Runemark et al., 2019; Taylor & Larson, 2019).

On a long-term evolutionary scale, ancient hybridization events between lineages are widespread in the evolutionary history of many taxa, resulting in diverse outcomes of adaptive radiation (Meier et al., 2017), adaptive introgression (Suarez-Gonzalez et al., 2016), hybrid speciation (Elgvin et al., 2017), or speciation reversal (Seehausen et al., 2008). In contrast to these at least partially stabilized hybrid populations, genomes in contemporary hybrid zones are in constant flux, and hybrid composition differs substantially among populations and individuals (McFarlane et al., 2022). We can study these different hybridization events to understand the early stages of hybridization and how novel hybrid zones evolve (Beninde et al., 2018; Coulter et al., 2020; Jahner et al., 2021; McFarlane et al., 2021). In addition, contemporary hybridization events can also provide insight into long-term genomic patterns within a clade. For example, Chaturvedi et al. (2020) showed that contemporary and ancient hybrid populations show similar patterns of introgression across the genome and that evolutionary outcomes can be predicted from contemporary hybrids in *Lycaeides* butterflies. Nouhaud et al. (2022) found that these processes can occur rapidly over short timescales and are driven by recombination rate variation and genetic load.

We studied the emergence and development of hybrid zones in the genus *Daphnia* (Crustacea, Cladocera), which are keystone species in freshwater ecosystems worldwide and are widely used as genomic and ecological model organisms (Ebert, 2022). Hybridization and introgression are common in the genus, and the geographic and ecological isolating barriers between closely related species have been studied extensively (Chin & Cristescu, 2021).

The life cycle of *Daphnia* makes them ideally suited for studying the history of hybridization events; they usually reproduce asexually with clonal individuals and switch to sexual reproduction with resting eggs under stressful environmental conditions (Ebert, 2005). These resting eggs can remain viable for decades and play a vital role in the temporal and spatial dispersal of *Daphnia* (Pietrzak & Ślusarczyk, 2006). Most resting eggs that are not reached by hatching stimuli, such as changes in light or temperature conditions, are eventually buried in lake sediments and form a resting egg bank (Brendonck & De Meester, 2003). In undisturbed sediments, the resting egg bank remains relatively stable long-term, with only eggs from the

surface layers that can be reached by hatching stimuli contributing to the genetic diversity of extant populations (Cáceres, 1998). By dating the sediment, these local biological archives can be used to hatch historical *Daphnia* genotypes or establish time-series genetic data from resting eggs to directly study adaptive responses and evolutionary changes (Brede et al., 2009; Frisch et al., 2014; Orsini et al., 2013). Recently, novel methods have been developed to address issues with low DNA yield and contamination that enable whole-genome sequencing of single historical resting eggs from the resting egg bank (Nickel & Cordellier, 2022; O'Grady et al., 2022).

Hybridization as a result of environmental change has been particularly well studied in the *Daphnia longispina* species complex (DLSC), which inhabits freshwater ecosystems in the northern hemisphere (Petrusek, Hobæk, et al., 2008). The species *D. galeata*, *D. longispina,* and *D. cucullata* can occur in the same habitats, but are separated by ecological gradients (Petrusek, Seda, et al., 2008), such as trophic level (Keller et al., 2008; Yin et al., 2014) and predation pressure (Spaak & Hoekstra, 1997), which allow for habitat niche separation and mosaic hybrid zones with complex hybridization dynamics. Previous studies identified hybrids between the species pairs using low-resolution genetic markers (Brede et al., 2009; Yin et al., 2014). Studies of hybrid zones often examine only two species, and multispecies hybridization, as found in *Populus* (Chhatre et al., 2018) and some birds (Natola et al., 2022; Ottenburghs, 2019), is rarely considered.

Recent hybridization can also be affected by habitat disturbances caused by human activities, which can result in new contact zones where closely related species meet that were previously fully or partially isolated due to their ecological divergence. Anthropogenic habitat disturbances that enable the coexistence of closely related species can increase hybridization rates (Grabenstein & Taylor, 2018) or result in the geographic shift of existing hybrid zones (Aguillon & Rohwer, 2022; Ryan et al., 2018). Hybridization in aquatic ecosystems has been impacted by large-scale environmental changes, such as climate change (Canestrelli et al., 2017), or local environmental changes, such as eutrophication (Alric et al., 2016) and the introduction of non-native species (Blackwell et al., 2021; Le Moan et al., 2021). This is well documented for *Daphnia* in peri-Alpine lakes, where increased eutrophication levels led to the successful establishment of *D. galeata*, which subsequently hybridized with the native *D. longispina* populations (Alric et al., 2016; Brede et al., 2009). The ability to utilize the resting egg bank and the unique features of aquatic hybrid zones make the DLSC very well suited to study the causes and consequences of recent hybridization over multiple generations and reconstruct admixture history.

In a previous study, we provided a *D. galeata* genome assembly and showed that hybridization in the DLSC is pervasive and shaped by multiple generations of hybridization and backcrossing, with introgression patterns varying among individuals and across the genome (Nickel et al., 2021). This suggests that hybridization in these populations was not limited to a single hybridization event and may be an ongoing process. For this study, we generated an improved chromosome-scale assembly with chromatin conformation capture sequencing to gain a deeper understanding of introgression patterns along each chromosome and reveal the extent and timing of hybridization events. Here, we focused on two populations in northern Germany that have been affected by eutrophication and were found to have extremely high levels of hybridization. To untangle the history of hybridization in these populations, we used broader time-series sampling and whole-genome data. Obtaining genomic time-series data to study genomic variation is difficult for many organisms and can often only be achieved using museum samples (Pinsky et al., 2021) or extensive sampling over shorter time periods (Gompert et al., 2021; Valencia-Montoya et al., 2020). The *Daphnia* resting egg bank enables us to obtain genomic data from different time points and stages of hybridization over several decades. We combined this approach with novel methods building on junction theory, which is based on how recombination breaks up genomic blocks of contiguous ancestry over time, and tracking markers along the genome to analyze the accumulation of delineations between these blocks, also known as junctions (Janzen et al., 2018). This allows more precise dating of recent admixture events (Janzen & Miró Pina, 2022) and detailed dissection of the different evolutionary histories that make up a hybrid's genome in the *Daphnia longispina* species complex.

Materials and Methods

Hi-C library preparation and sequencing

Several hundred individuals of the M5 clone, used to generate the previous *D. galeata* genome assembly, were raised in bulk and fed ad libitum with the algae *Acutodesmus obliquus*. After collection and water removal, the organisms were rinsed with PBS buffer, pelleted through gentle centrifugation, and the buffer was removed. The sample was then resuspended in 1 ml of 1% formaldehyde solution (in PBS). After 20 min of incubation at room temperature with periodic mixing, glycine was added to a final concentration of 125 mM. The sample was then incubated at room temperature for 15 min, pelleted through gentle centrifugation, rinsed with PBS, pelleted again, and the supernatant was removed. The sample was resuspended in PBS and ground with a loose-fitting Dounce homogenizer (ten strokes). The homogenized sample was pelleted by spinning down at 6000 × g for 5 min, the supernatant was removed, and the pellet was snap frozen. Library preparation with the Proximo Hi-C Animal Kit and sequencing

on an Illumina HiSeq 4000 to generate 150 bp paired-end reads were conducted by an external company (Phase Genomics, Seattle, WA).

Hi-C scaffolding

Chromosome-scale scaffolding was performed with the *D. galeata* assembly (GenBank accession number: GCA_918697745.1) from Nickel et al. (2021) and the Hi-C sequencing data as input, using Juicer v1.6 (Durand, Shamim, et al., 2016) in combination with BWA v0.7.17 (Li, 2013) and 3D-DNA (commit 529ccf4; Dudchenko et al. 2017) with default parameters. The resulting scaffolds were manually curated with Juicebox v1.11.08 (Durand, Robinson, et al., 2016) and 3D-DNA was executed again to create the final scaffold set. Basic assembly contiguity metrics were calculated with Quast v5.1.0rc1 (Mikheenko et al., 2018). This reference genome was used for all subsequent analyses presented here.

To annotate protein-coding genes in the assembled scaffolds, GeMoMa v1.8 (Keilwagen et al., 2016, 2018), a reference-based method, was used. Paired and unpaired RNAseq reads (Huylmans et al., 2016) were mapped against the chromosome-scale assembly with HISAT v2.2.1 (Kim et al., 2019) and subsequently merged and sorted with samtools v1.13 (Danecek et al., 2021) to create a single BAM file. Together with the previously published assembly of *D. galeata* (g) and its annotation in gff format (a) as reference, the chromosome-scale assembly as target (t) and the mapping of the RNAseq reads described above (ERE.m; r=MAPPED) were used as input for the GeMoMa CLI GeMoMaPipeline along with the parameters o=true AnnotationFinalizer.r=NO.

Sample collection

Multiple sediment cores were collected with a gravity corer (UWITEC, Mondsee, AT) from the mesotrophic lake Selenter See (SE) and the eutrophic lake Eichbaumsee (EIC), Germany, in September 2018 and 2020, respectively. Cores were collected from the deepest area of each lake to minimize the effects of bioturbation and previous disturbance of the sediment layers (Table S4.1). Cores (90 mm diameter) were segmented horizontally into 1 cm thick sections and stored at 4 °C in the dark to prevent any hatching of resting eggs. For Selenter See, layers of a single core were freeze-dried to determine the sedimentation rate using radioisotope dating (^{137}Cs and ^{210}Pb) and to estimate the age of the resting eggs from each layer. This was not done for Eichbaumsee as it is an excavation lake that was constructed in 1976 and is not connected to other water bodies. At the bottom of the cores, we found a light-grayish sand layer as described by Förster et al. (2021), indicating that the cores capture the entire history of the lake.

Additional resting eggs and living *Daphnia* from the water column were collected from Eichbaumsee during a previous study (Nickel et al., 2021). Firstly, lake sediment was collected by hand from the shoreline of Eichbaumsee and stored at 4 °C. Secondly, zooplankton samples were collected by hauling a 150 µm plankton net through the water column from which six *Daphnia* clonal lines were established in the laboratory.

Single genotypes generated in a previous study (Nickel et al., 2021) representing the three parental species *D. galeata*, *D. longispina*, and *D. cucullata* from various European lakes were selected and are used here as “parental species genotypes”. The clonal lines were either established from individuals sampled from the water column or hatched from a resting egg and were maintained as monoclonal cultures in the laboratory through asexual reproduction. The species identity of these genotypes was determined by a combination of morphology, mitochondrial sequences, and nuclear markers. This panel included five genotypes for *D. galeata*, four for *D. longispina*, and three for *D. cucullata*. Additional information on the parental species genotypes, DNA isolation, and whole-genome sequencing can be found in Nickel et al. (2021). We generated additional resequencing data for the *D. longispina* clonal line KL11, which had low coverage in the previous reference panel.

Whole genome amplification of resting eggs and DNA extraction from *Daphnia* cultures

Collection of *Daphnia* ephippia from sediment and isolation of the resting eggs followed by whole genome amplification (WGA) was performed as described by Nickel et al. (2021). Briefly, *Daphnia* eggs were collected from each sediment layer by resuspending small amounts of sediment in ddH_20, eye-spotting ephippia, and storing the samples at −20 °C in the dark. Then, the ephippial casings were opened with insect needles and tweezers to isolate intact resting eggs. The isolated resting eggs were stored in 3 µl 1× PBS at −80 °C. WGA of single eggs was performed with the REPLI-g Mini Kit (Qiagen) following the manufacturer’s protocol for the amplification of genomic DNA from blood or cells. The work was conducted in a PCR Workstation Pro (VWR) and surfaces and tools were treated with UV light and DNA-ExitusPlus (PanReac AppliChem) to reduce the risk of contamination. The amplified product was quantified with a Qubit Fluorometer (Thermo Fisher Scientific).

As WGA from starting material of limited quantity and/or quality may result in failed amplification of the source DNA or contamination from other organisms, a pre-sequencing screening was performed by amplifying mitochondrial markers to identify *Daphnia* and bacterial DNA in the amplified product as described by Nickel and Cordellier (2022). Only samples with a successful amplification of the *Daphnia* 16S fragment and low or no amplification of the bacterial 16S fragment, indicating low bacterial contamination, were used

for further sequencing steps. Based on these criteria, the successful amplification products were purified with 0.4× Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove small fragments and eluted in 60 µl 1× TE buffer.

For each of the six EIC clonal lines and the reference clonal line KL11, 20 adult *Daphnia* were pooled to obtain sufficient amounts of DNA, and total genomic DNA was isolated using a modified CTAB extraction method (Cristescu et al., 2006).

Library preparation and sequencing

Quantification and quality control of the genomic DNA samples were performed using Qubit and Nanodrop instruments (Thermo Fisher Scientific) and libraries were prepared either in-house with the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs) or at the sequencing company Novogene (Cambridge, UK). All libraries were sequenced using 150 bp paired-end reads on an Illumina NovaSeq 6000 at Novogene. Information on all samples, the specific library preparation protocol, and ENA accession numbers is provided in Table S4.2.

Sequencing QC, trimming, and contamination screening

The quality of raw reads was assessed with FastQC v0.11.7 (Andrews, 2010). Where necessary, poly-G tails that can appear in NovaSeq reads, were trimmed with fastp v20.0 (Chen et al., 2018). We trimmed Illumina adapters, removed base calls at the end of reads with a Phred quality score <20 and reads with a quality score <15 across a 4 bp sliding window, and discarded reads with a length <70 bp using Trimmomatic v0.38 (Bolger et al., 2014).

As WGA of single resting eggs can result in little or no amplification of *Daphnia* DNA and amplification of DNA contamination from other organisms, we examined contamination in the WGA samples with FastQ Screen v0.14.0 with the BWA mapping option and default parameters. Trimmed reads were mapped against a custom database containing the possible laboratory contaminants *Homo sapiens* (GCA_000001405.22), the UniVec database (https://ftp.ncbi.nlm.nih.gov/pub/UniVec/), a bacterial (https://ftp.ncbi.nlm.nih.gov/refseq/release/bacteria/) and a viral reference set (https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/), all downloaded in April 2018, as well as the *D. galeata* genome (GCA_918697745.1) and the draft genome of the food algae *Acutodesmus obliquus* (GCA_002149895.1). Samples with <25% reads mapped to the *D. galeata* genome or >20% contamination with one or multiple other genomes were excluded from further analysis.

Read alignment and variant calling

The trimmed reads were mapped to the improved chromosome-scale *D. galeata* reference genome (available in Zenodo, doi:10.5281/zenodo.8069314) with BWA-MEM v0.7.17 using

the -M parameter to allow for Picard compatibility, and read group identifiers were added to combine samples for which multiple libraries or runs were sequenced (Li & Durbin, 2009). PCR duplicates were marked and removed from the BAM files using Picard v2.25.4 (http://broadinstitute.github.io/picard/).

We combined two variant calling approaches to obtain high-quality SNPs. The first analysis was performed within the Genome Analysis Toolkit (GATK v4.2.2.0; McKenna et al., 2010) and adapted from the GATK best practice recommendations (Van der Auwera et al., 2013). Variants for each sample were called with GATK HaplotypeCaller and the --emitRefConfidence GVCF option, resulting in a genomic variant call format (gVCF) file with information on each position for each individual (Poplin et al., 2018). All gVCF files were merged using CombineGVCFs, and joint genotyping was performed with GenotypeGVCFs. Only SNPs were selected from the resulting VCF file and hard filtering values were applied using VariantFiltration to remove SNPs with a QualByDepth <10, StrandOddsRatio >3, FisherStrand >60, mapping quality <40, MappingQualityRankSumTest <−8, and ReadPosRankSumTest <−5.

Further, we estimated genotype likelihoods implemented in ANGSD v0.931-2 (Korneliussen et al., 2014), which is designed for low-coverage sequence data. SNPs were inferred based on genotype likelihoods (-GL 2 -doMajorMinor 1 -doMaf 1) using a threshold likelihood ratio for SNP calling (-SNP_pval 1e−06) and the following quality filters: -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMinDepth 400 -setMaxDepth 2000 -minMaf 0.05 -doCounts 1.

Subsequently, only biallelic SNPs that were identified by both GATK and ANGSD and for which genotypes were available for at least 80% of the individuals were selected using VCFtools v0.1.5 (Danecek, et al. 2011). The final SNP data set for downstream analyses contained 607,440 SNPs across the 68 samples.

Ancestry and population structure to identify hybrid individuals

We conducted principal component analysis (PCA) to visualize the clustering of all genotypes in R v4.1.0 (R Core Team, 2021) with the package SNPRelate v1.20.1 (Zheng et al., 2012). We pruned the SNP data set for linkage disequilibrium (r^2 >0.5) in a sliding window using the snpgdsLDpruning function before conducting the PCA using the snpgdsPCA function with default settings for all biallelic sites.

To evaluate individual genetic admixture in the data, the SNP VCF file was converted to BED format using PLINK v1.9 (Chang et al., 2015). The cross-validation error was estimated for genetic clusters (*K*) of ancestral populations from 1 to 5 using ADMIXTURE v1.3.0 and

admixture analysis was performed for the most appropriate $K = 3$ value with 10-fold cross-validation (Alexander & Lange, 2011).

The ADMIXTURE analysis suggested that the population EIC had a history of *D. galeata* × *D. longispina* hybridization, and the population SE had a history of both *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata* hybridization. To identify fixed sites in the parental species pairs, we used the "ancestry painting" approach described by Barth et al. (2020) and Runemark et al. (2018), which classifies sites according to their F_{ST} value using the previously defined parental genotypes. We considered a site to be fixed if it was present in all reference parental species and a specific allele was fixed in all individuals belonging to one parental species and a different allele in the other parental species. To plot ancestry paintings, sites were identified that showed at least 90% of genotype completeness and SNPs were thinned to be at least 100 bp apart using available ruby scripts (https://github.com/mmatschiner/tutorials/tree/master/analysis_of_introgression_with_snp_data, last accessed April 2022) to show the ancestry of hybrid individuals and visualize regions potentially affected by introgression.

For some of the samples, mitochondrial genome assemblies used to reconstruct phylogenetic relationships were already available (Nickel et al., 2021). For all other potential hybrid individuals, de novo assembly of the mitochondrial genomes was conducted using the "de novo assembly" and "find mitochondrial scaffold" modules implemented in MitoZ v2.4 (Meng et al., 2019). To assign each mitochondrial genome to a maternal species, it was searched against a custom database built with all the mitochondrial genomes constructed for the parental species genotypes using BlastN v2.12.0+ (Camacho et al., 2009).

Estimating time since hybridization event

To accurately date recent admixture events in hybrid populations, we used the "junctions" approach (Janzen et al., 2018). This approach is an extension of previous junction theory approaches that, for example, were able to classify up to seven generations of hybridization in ducks using ddRAD-seq data (Lavretsky et al., 2019).

We performed junctions analysis on two distinct data sets to determine when the hybridization event of each individual took place; one collected from population EIC and one collected from population SE. The reference samples were labeled as *D. galeata, D. longispina*, or *D. cucullata,* and all population samples as the species pair *D. galeata* × *D. longispina* or *D. galeata* × *D. cucullata* (Table 4.1). In population EIC, we found only *D. galeata* and *D. longispina* ancestry in the ADMIXTURE analysis, making the *D. galeata* × *D. longispina* species pair the most probable. For the assignment of individuals in population SE to the species pair *D. galeata* × *D. longispina* or *D. galeata* × *D. cucullata*, we used ADMIXTURE results when possible or their position in the PCA.

Using the fixed sites obtained from the ancestry painting for *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata*, we had 129,442 and 173,672 sites, respectively. Then, we transformed the fixed sites from VCF format into Ancestry HMM (Corbett-Detig & Nielsen, 2017) input format using a custom Python script (available in Zenodo, doi:10.5281/zenodo.8069314). Briefly, the script processes the entire genome on a per-SNP basis and records allele counts in both the parental species and the hybrids.

Then, we used Ancestry HMM v1.0 (Corbett-Detig & Nielsen, 2017) to infer local ancestry per individual using two different recombination rates: 6.78 cM/kb obtained from *D. magna* (Dukić et al., 2016) and 7.57 cM/kb obtained from female *D. pulicaria* (Wersebe et al., 2023). We increased the power of Ancestry HMM by jointly analyzing samples known to stem from similar sampling efforts (based on where the individuals were collected). Population SE was analyzed as a single group ("SE core"), and population EIC was separated into five groups; animals collected from the water column ("clones"), undated sediment ("sediment–top"), and three sections of the sediment core ("core–top", "core–middle", "core–bottom") (Table 4.1).

Ancestry HMM provides a posterior probability of local ancestry at each SNP (e.g., of observing complete *D. longispina* ancestry, complete *D. galeata* ancestry, or heterozygous ancestry). We translated these posterior probabilities into hard ancestry calls depending on a cut-off threshold and varied the threshold in steps $1-10^{[-6,\,-5,\,-4,\,-3,\,-2,\,-1]}$. To infer time since admixture estimates on a per-sample basis, the obtained hard ancestry calls were then analyzed using the junctions package v2.0.3 in R (Janzen, 2021; Janzen & Miró Pina, 2022), which takes local ancestry as input. We used the function "estimate_time_diploid" using analysis_type "individuals", in order to obtain age estimates split per individual, assuming that all scaffolds belonging to the same individual have a shared genomic history. The analysis was performed assuming a large population size of 1,000,000 individuals and equal ancestry, such that initial heterozygosity was 0.5. To obtain per-chromosome age estimates, we repeated the analysis, but now used analysis_type "chromosomes", which no longer assumes that all scaffolds belonging to the same individual have a shared history, allowing for independent age inference on a per-chromosome basis. For visualization, the unplaced scaffolds were removed and the local ancestry and position of junctions were plotted for the ten chromosome-scale scaffolds.

Results

Chromosome-scale Hi-C genome assembly

In total, 213,164,294 Hi-C read pairs were generated, which provided a uniform distribution of around 50× coverage of the *D. galeata* genome. Scaffolding with Juicer and 3D-DNA resulted in a higher number of contigs and scaffolds compared to the published assembly, showing a significant amount of corrections in the form of splits. After manual curation, the number of

contigs and scaffolds increased even more, mainly because additional misassemblies that were not detected by the algorithms could be corrected. With 68 Mb, the scaffold N50 of the uncurated assembly is unreasonably high compared to other chromosome-scale assemblies of the genus *Daphnia* (Barnard-Kubow et al., 2022; Lee et al., 2019; Wersebe et al., 2023). This is due to false scaffolding that combined sequences from five and three different chromosomes into one scaffold, respectively (Figure S4.1A). After manually correcting these macroscopic errors, as well as many smaller assembly errors (presumably originating from scaffolding with PacBio reads of the published assembly from Nickel et al. (2021)), ten large scaffolds representing the assumed karyotype for *D. galeata* can be seen (Figure S4.1B; Beaton & Hebert, 1994). The ten largest scaffolds contain more than 96.6% of the total length of the assembly. Correcting misassemblies lowered the contig N50 after manual curation compared to the previously published *D. galeata* assembly. However, the scaffold N50 of 12.4 Mb of the manually curated assembly was improved and similar to published chromosome-scale assemblies in the genus *Daphnia* (Table S4.3).

GeMoMa yielded 16,438 genes and 120,347 coding sequences with a total length of 46,528,337 and 25,530,649 bp, respectively. The median lengths of genes and CDSs are 1,807 and 155. The mean number of CDSs per gene is 6.6. Searching for single copy orthologs of Arthropoda (arthropoda_odb10) with BUSCO v5.2.2 (Manni et al., 2021) resulted in 90.8% completeness (C: 90.8% [S: 89.0%, D: 1.8%], F: 0.9%, M: 8.3%, n: 1,013).

Time-series sampling and genotyping

Resting eggs were isolated from four different depth sections of sediment cores collected from Selenter See (SE), and whole-genome resequencing was carried out for 17 WGA samples that passed the PCR contamination screening. For the population SE, the youngest eggs analyzed originated from the 0–1 cm section of the core, which corresponds to the year ~2016, and the oldest eggs from the 4–5 cm section correspond to the year ~1992 (Andersen T, personal communication). After screening the sequences for contamination and removing data with only very few reads mapping to the *D. galeata* reference genome, 14 individual genotypes from resting eggs remained for further analysis (“SE core”, Table S4.2).

For the population from Eichbaumsee (EIC), we generated whole-genome sequences for six clonal lines sampled from the water column. Then, whole-genome resequencing data was generated for 59 WGA samples (Table S4.2). After removing contaminated sequences, 36 individuals remained for analysis. Eight resting eggs were collected directly from the shoreline and represent the most recent eggs, and 28 resting eggs originated from sediment cores.

The EIC sediment cores were not dated, but the lake was artificially created in 1976 and the sediment cores represent a complete time series from the time of sampling (2020) to the

formation of the lake, going back five decades. To allow comparison between ages in the genomic analysis, we pooled eggs from adjacent sections in the sediment core into three sample groups that represent the youngest eggs ("core–top", 0–5 cm, 9 total), eggs of intermediate age ("core–middle", 13–18 cm, 12 total), and the oldest eggs ("core–bottom", 24–29 cm, 7 total), as well as one group each from clones collected in the water column ("clones", 6 total) and recent eggs of unknown age ("sediment–top", 8 total) (Table S4.2). We used whole-genome data from 12 clonal lines with previously established species identities generated by Nickel et al. (2021) as reference data representing the three parental species and their diversity.

We analyzed 1.5 billion mapped sequencing reads and obtained 607,440 biallelic SNPs for 68 individuals using an approach that combined hard-called genotypes and SNPs based on genotype likelihoods. The average genome coverage per individual was 16.86× (range: 1.73–95.37×) for the population samples and 50.9× (range: 23.18–69.93×) for the reference samples (Table S4.4).

Ancestry and population structure to identify hybrid individuals

The ADMIXTURE analysis strongly supported a *K* value of 3, which is biologically meaningful as it clearly differentiates the three known parental species into separate clusters (Figure 4.1). Although ADMIXTURE is not always reliable and may fail to identify hybrids, especially those with highly asymmetric parental contributions (Kong & Kubatko, 2021), we consider this sufficient to identify potential hybrids for further analysis. In population EIC, we detect *D. galeata* (19) and *D. longispina* (9) and 14 hybrid individuals that show admixture between the two species. The first appearance of sampled hybrid individuals in this lake is recent, as we could only detect them in the "core–top" section. For population SE we find five *D. galeata* and admixture between *D. galeata* and *D. cucullata* (6) and *D. galeata* and *D. longispina* (3). The number of samples from the different time periods is small, but hybrids still appear in each sampled section.

The method used to estimate time since admixture requires a previously known species pair (Janzen & Miró Pina, 2022), but can detect older admixture events that are potentially not identified by ADMIXTURE. Therefore, we classified all individuals as belonging to the *D. galeata* × *D. longispina* or *D. galeata* × *D. cucullata* species pair based on ADMIXTURE and PCA results. The ADMIXTURE analysis detected a total of 23 hybrid individuals and their respective species pair. When individuals appeared unadmixed in the ADMIXTURE analysis, the PCA results were used to classify species pairs. The PCA showed that the three parental species are grouped into distinct clusters, while population EIC is spread between the *D. galeata* and *D. longispina* clusters, and population SE mostly between the *D. galeata* and *D.*

cucullata clusters, both with no clear temporal substructure (Figure S4.2). Based on the position in the PCA plot, the remaining individuals were classified as belonging to one of the species pairs (Table 4.1).

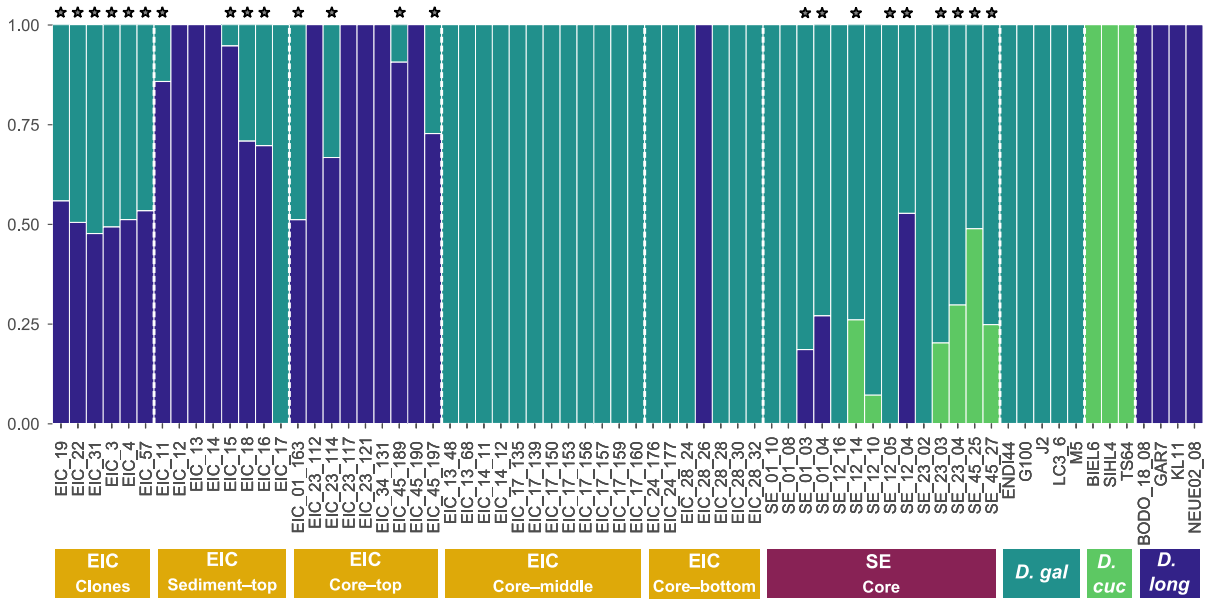


Figure 4.1. Ancestry proportions based on ADMIXTURE analysis ($K = 3$) admixture plot to identify potential hybrids. Population EIC and SE are ordered by sample age, and the classified hybrid individuals are marked with a star. Clusters are colored according to the reference species *D. galeata* (teal), *D. cucullata* (green), and *D. longispina* (dark blue).

To visualize ancestry and introgression across the genome, we generated ancestry paintings, with multiple breakpoints along a chromosome indicating multiple generations of recombination. In general, we found many chromosome pairs where the majority of sites carry ancestry from both species, which we refer to as heterospecific regions, in the ancestry painting that uses only ancestry informative fixed sites.

We analyzed a total of 16 *D. galeata* × *D. longispina* hybrids found across both populations, all of which carry *D. galeata* mitochondria (Figure 4.2A). Reference sample KL11 and potential hybrid individual EIC_23_114 were removed from the analysis due to the high proportion of missing sites in the samples. The six samples that originated from clonal lines and therefore represent animals living in the water column at the time of sampling show extremely similar ancestry patterns (mostly heterospecific chromosome pairs with small regions of homospecific *galeata* ancestry on some chromosome pairs, e.g., 6, 7, and 10). Compared to *D. galeata* × *D. cucullata* hybrids, the ancestry tract length is significantly shorter, with many breakpoints between regions of *galeata* and *longispina* ancestry on a single chromosome leading to more heterospecific regions. However, we also find individuals with some chromosome pairs that have not recombined and thus show pure *galeata* or *longispina* ancestry (e.g., sample SE_01_04 and EIC_01_163).

We analyzed six *D. galeata* × *D. cucullata* hybrids, all found in population SE (Figure 4.2B). We find pure *galeata* chromosome pairs or homospecific regions in most samples, e.g., chromosomes 3 and 5. Pure *cucullata* ancestry in some regions, e.g., chromosomes 2 and 3, is mostly found in sample SE_45_25. In addition, we also find chromosome pairs with homospecific *galeata* regions and heterospecific regions caused by a recombination breakpoint on one of the homologous chromosomes, e.g., chromosomes 2 and 4. Samples SE_12_10 and SE_12_14 show ancestry from both species but mostly homospecific sites. All but one of the samples carry *D. cucullata* mitochondria.

Using time-series genomic data and time since admixture estimates to understand hybridization history

We inferred local ancestry using Ancestry HMM with all fixed sites identified in the ancestry paintings as ancestry informative markers, and based on this we inferred the time since admixture for each individual using junctions to obtain a detailed understanding of the hybridization history. This was done on an individual level and for each specific sample group to take into account that each group originates from a different time period and the time since admixture is different for each. Estimates of time since admixture in years are based on the assumption that *Daphnia longispina* species undergo one sexual reproduction per year (Jankowski & Straile, 2004; Spaak et al., 2004).

The threshold used for hard ancestry calls had a large effect on the number of remaining SNPs (Table S4.5). Using very strict thresholds ($1-10^{[-6,-5,-4]}$), several samples had no or very few remaining SNPs. We chose a threshold of $1-10^{-2}$ to keep a total of 123,003 sites for analysis for the species pair *D. galeata* × *D. longispina* (12,300 average per chromosome; 95% CI: [6,474, 26,595]) and 164,741 sites for the species pair *D. galeata* × *D. cucullata* (16,474 average per chromosome; 95% CI: [8,899, 30,932]). However, four samples were removed due to a high amount of missing data.

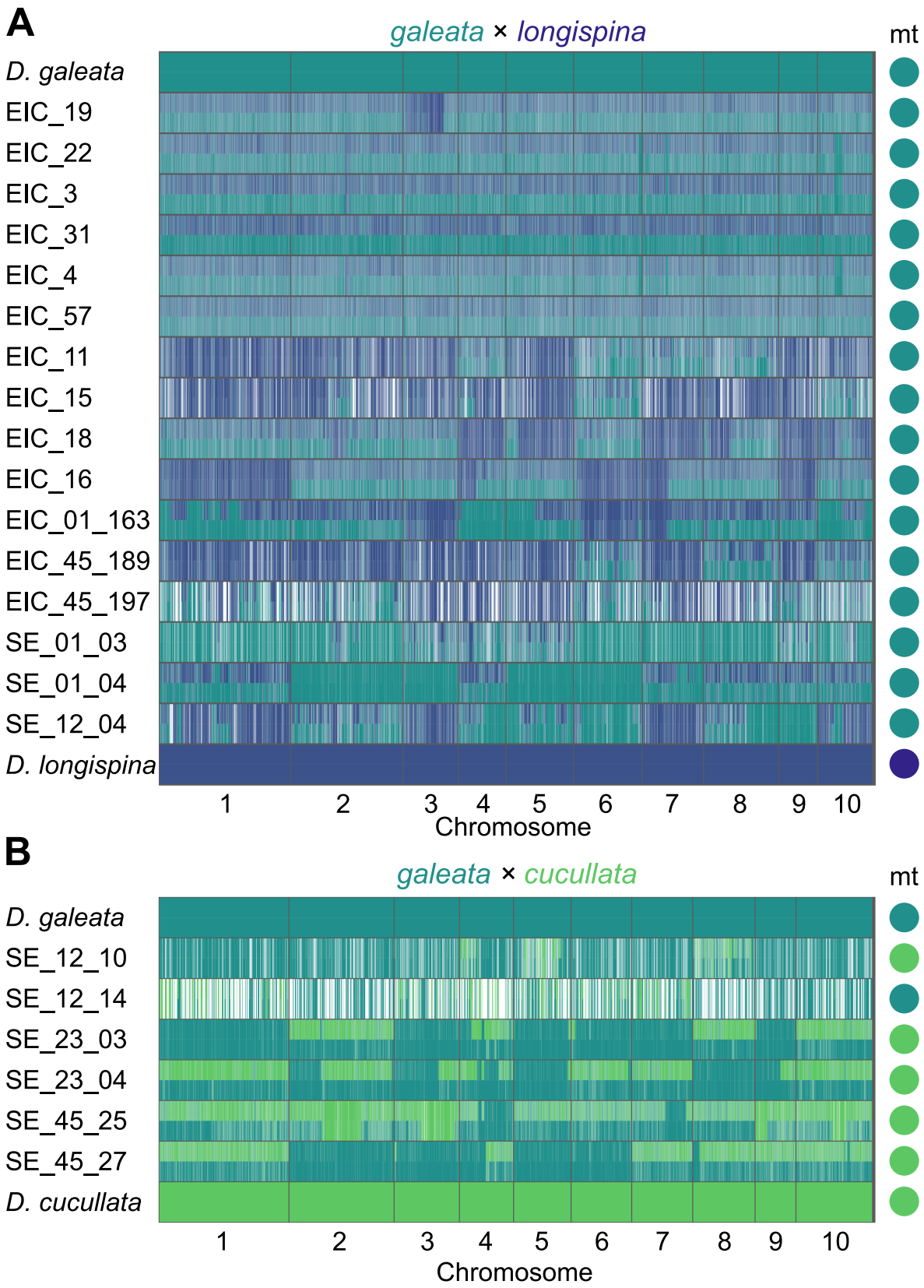


Figure 4.2. Ancestry painting of hybrid individuals identified by the ADMIXTURE analysis along the ten chromosomes for (A) *D. galeata* × *D. cucullata* for 80,075 sites and (B) *D. galeata* × *D. longispina* for 28,019 sites. Each row represents one individual, with teal indicating alleles inherited from *D. galeata*, dark blue from *D. longispina,* and green from *D. cucullata*. Sites with two colors in the same row are considered heterospecific. For better visibility, the five *D. galeata* and three *D. cucullata* and *D. longispina* reference samples with completely fixed sites were collapsed into a single row each.

Because no recombination rate is available for a species in the DLSC or a closely related species, and because the analysis is sensitive to recombination rate assumptions, we tested the analysis with two different recombination rates obtained from *D. magna* and *D. pulicaria* to assess the most suitable one (Figure S4.3; Janzen & Miró Pina, 2022). Using the recombination rate of 7.57 cM/kb from female *D. pulicaria* (Wersebe et al., 2023), we obtained a mean time since admixture of 16.1 years for all 52 individuals (95% CI: [9.7, 23.6]). For the recombination rate of 6.78 cM/kb *D. magna* (Dukić et al., 2016), this estimate was slightly higher (20.1 years; 95% CI: [14.1, 26.7]). As the time since admixture estimates for both recombination rates for all samples gave very similar and stable results, we decided to use the recombination rate from *D. pulicaria*, which is more closely related to the DLSC than *D. magna* (Cornetti et al., 2019), and all further reported results use the recombination rate of 7.57 cM/kb.

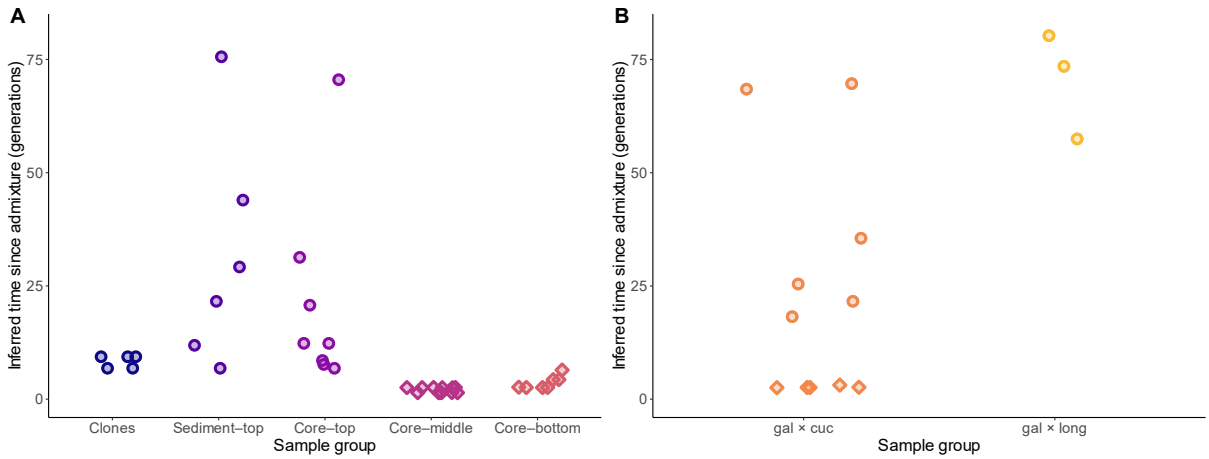


Figure 4.3. Inferred time since admixture for (A) the 38 individuals from the population EIC for the five sample groups and (B) the 14 individuals from the population SE for *D. galeata* × *D. cucullata* and *D. galeata* × *D. longispina* hybrids. Circles indicate that samples are correctly classified as admixed in the junctions analysis and diamonds that the samples are most likely unadmixed parental species.

The time since admixture for each individual from different periods in the lakes' histories gives us insight into how the populations developed (Table 4.1). For population EIC, all individuals in the "clones" group were found to have a time since admixture estimate of 6.8 or 9.4 years (Figure 4.3A). This is supported by the very similar ancestry patterns found for this group in the ancestry paintings. In both the "sediment–top" and "core–top" groups, we find extremely variable time since admixture estimates ranging from 6.8 up to 75.6 years. The "core–middle" and "core–bottom" groups are characterized by low time since admixture estimates for all individuals. However, this is driven by very small genomic regions of changing ancestry, and only few junctions are found across the genome (Figure S4.6, Figure S4.8). Based on these low estimates of time since admixture and the genomic patterns, we suggest that these individuals are most likely not recently admixed and belong to one of the parental species, which is also consistent with the ADMIXTURE results for these individuals. We believe that

these junctions could be caused by sites that are not completely fixed in the parental reference species that were used to determine the fixed sites.

In population SE two different species pairs were analyzed (Figure 4.3B). The three *D. galeata* × *D. longispina* hybrids all gave similarly high time since admixture estimates (57.5–80.2 years). For *D. galeata* × *D. cucullata* hybrids, we find five individuals with very low estimates which indicate that they are most likely parental *D. galeata*. The estimates for the other nine individuals range from 18.2 to 69.7 years. We do not have enough samples per section to compare time since admixture patterns between them.

Table 4.1. Time since admixture results for all individuals using the recombination rate of 7.57 cM/kb and a hard ancestry call threshold of $1-10^{-2}$.

Sample name	Species pair assignment	Group	Time since admixture (years)
EIC_19	gal × long	Clones	9.4
EIC_22	gal × long	Clones	9.4
EIC_3	gal × long	Clones	6.8
EIC_31	gal × long	Clones	6.8
EIC_4	gal × long	Clones	9.4
EIC_57	gal × long	/	/
EIC_11	gal × long	Sediment–top	11.9
EIC_12	gal × long	Sediment–top	6.8
EIC_13	gal × long	Sediment–top	29.2
EIC_14	gal × long	Sediment–top	75.6
EIC_15	gal × long	Sediment–top	21.6
EIC_18	gal × long	/	/
EIC_16	gal × long	Sediment–top	44
EIC_17	gal × long	/	/
EIC_01_163	gal × long	Core–top	20.8
EIC_23_112	gal × long	Core–top	8.5
EIC_23_114	gal × long	/	/
EIC_23_117	gal × long	Core–top	6.8
EIC_23_121	gal × long	Core–top	12.3
EIC_34_131	gal × long	Core–top	12.3
EIC_45_189	gal × long	Core–top	31.3
EIC_45_190	gal × long	Core–top	7.7
EIC_45_197	gal × long	Core–top	70.5
EIC_13_48	gal × long	Core–middle	2.5
EIC_13_68	gal × long	Core–middle	1.4
EIC_14_11	gal × long	Core–middle	2.6
EIC_14_12	gal × long	Core–middle	2.6
EIC_17_135	gal × long	Core–middle	2.6
EIC_17_139	gal × long	Core–middle	1.4
EIC_17_150	gal × long	Core–middle	1.4
EIC_17_153	gal × long	Core–middle	1.4
EIC_17_156	gal × long	Core–middle	1.4
EIC_17_157	gal × long	Core–middle	1.4

EIC_17_159	gal × long	Core–middle	2.5
EIC_17_160	gal × long	Core–middle	2.6
EIC_24_176	gal × long	Core–bottom	2.6
EIC_24_177	gal × long	Core–bottom	2.6
EIC_28_24	gal × long	Core–bottom	2.6
EIC_28_26	gal × long	Core–bottom	6.4
EIC_28_28	gal × long	Core–bottom	4.3
EIC_28_30	gal × long	Core–bottom	2.6
EIC_28_32	gal × long	Core–bottom	4.3
SE_01_03	gal × long	SE core	80.2
SE_01_04	gal × long	SE core	57.7
SE_12_04	gal × long	SE core	73.5
SE_01_08	gal × cuc	SE core	2.6
SE_01_10	gal × cuc	SE core	3.1
SE_12_05	gal × cuc	SE core	21.6
SE_12_10	gal × cuc	SE core	2.6
SE_12_14	gal × cuc	SE core	18.2
SE_12_16	gal × cuc	SE core	69.7
SE_23_02	gal × cuc	SE core	2.6
SE_23_03	gal × cuc	SE core	25.4
SE_23_04	gal × cuc	SE core	35.5
SE_45_25	gal × cuc	SE core	2.6
SE_45_27	gal × cuc	SE core	68.4

We then estimated the time since admixture for each chromosome per individual for both populations, because the junction patterns showed large differences between chromosomes, which suggests that recombined chromosomes may have originated from different admixture events. The per-chromosome time since admixture estimates show a large variation among the chromosomes in the hybrid individuals, especially those with higher estimates, with a diverse ancestry ranging from more recently introduced chromosomes to chromosomes with much higher time since admixture estimates (>100 years) and higher heterozygosity for ancestry informative sites (Figure S4.4, Figure S4.5).

In addition, we analyzed how ancestry was distributed along the ten chromosomes and whether the hybrid individuals show similar patterns using the local ancestry probabilities and junction positions on which the time since admixture estimates were based. The five groups of population EIC show distinct patterns (Figure S4.6, Figure S4.8).

The EIC "clones" group shows an overall very similar number of junctions, corresponding to similar time since admixture estimates, with the junctions being evenly distributed at similar positions across all chromosomes in the five analyzed individuals. The local ancestry shows that the majority of genomic regions are heterospecific across chromosome pairs. Taken together, these results show similar but not completely identical genomic patterns, which could

occur if two or more identical clones were included in the data set, which is possible in lakes dominated by a particularly successful hybrid clone (Griebel et al., 2015).

For the EIC "sediment–top" group, the six analyzed individuals show a higher junction density on certain chromosomes (e.g., chromosomes 2 and 6). The local ancestry for each individual shows a mixture of entire chromosome pairs of *longispina* ancestry, chromosome pairs with a majority of *longispina* ancestry and smaller regions of *galeata* ancestry, and chromosome pairs with an extremely high number of junctions that delineate *galeata*, *longispina*, and heterospecific regions (especially for samples EIC_14 and EIC_16 with the highest time since admixture estimates). Overall, the EIC "core–top" group presents very similar patterns of junction distribution and local ancestry to the EIC "sediment–top" group, with slightly more individuals carrying entire chromosome pairs of *longispina* ancestry.

The EIC "core–middle" group has only very few total junctions across the genome, and they are often found in multiple individuals (e.g., chromosomes 1 and 6). This is also the case for the EIC "core–bottom" group, but not at exactly the same positions. The local ancestry for both EIC "core–middle" and EIC "core–bottom" also suggests that the majority of the genome of these individuals is pure *galeata* with very small regions of *longispina* or heterospecific ancestry that could be caused by not completely fixed sites or single chromosomes with traces of ancient introgression.

In population SE, two different species pairs were found and used for the junctions analysis (Figure S4.7, Figure S4.9). For *D. galeata* × *D. cucullata*, the five individuals, we assumed to be the parental species showed only very few putative junctions and consistent *galeata* ancestry. For all other hybrid individuals, the junction density varies between individuals and chromosomes, and the local ancestry shows more states with increasing age. For the three *D. galeata* × *D. longispina* individuals, we find a high junction density, but very different patterns for each individual, i.e., some chromosomes have only a few junctions while others have a high density. This appears to be similar to *D. galeata* × *D. longispina* from population EIC with similar time since admixture estimates (e.g., EIC_13). The local ancestry tracts are very short and show all possible states.

As an example, we show chromosome 1 for all *D. galeata* × *D. longispina* from both populations and chromosome 4 for *D. galeata* × *D. cucullata* from population SE to take a closer look at the distribution of junctions along one chromosome and identify potential similarities between individuals (Figure 4.4). Genomic regions without any introgression could point to the role of these regions in reproductive isolation between the species.

In all the *D. galeata* × *D. longispina* individuals, the junctions are distributed across the entire chromosome 1, with no major regions where none appear in at least one individual (Figure 4.4A). Some junctions appear to be fixed in most individuals, e.g., at the beginning of the chromosome. The junctions in all *D. galeata* × *D. cucullata* individuals are evenly distributed across chromosome 4, with the exception of a lack of junctions from approximately 12 Mb to the end of the chromosome (Figure 4.4B). Overall, junction patterns are rarely shared among individuals, even for samples with similar time since admixture estimates.

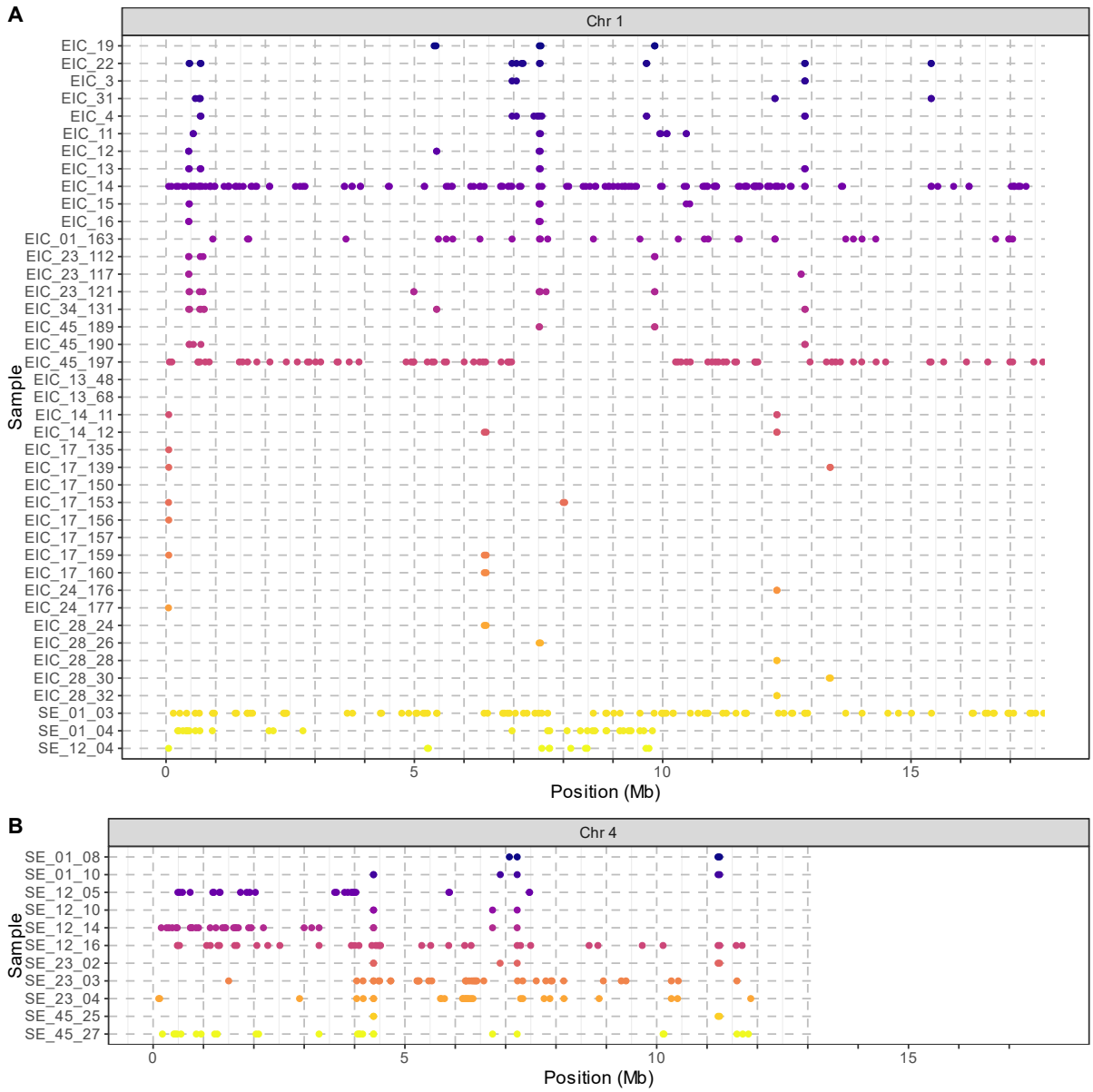


Figure 4.4. Junction positions (A) on chromosome 1 for all *D. galeata* × *D. longispina* individuals from populations EIC and SE, and (B) on chromosome 4 for all *D. galeata* × *D. longispina* hybrids from population SE.

Discussion

Using time-series genomic data, we aimed to understand the extensive hybridization history in the DLSC. Our results reveal a complex history of hybridization among three different species, support recurrent hybridization in different populations and between different species over several decades, and show high variation in hybrid ancestry among individuals and across the genome. In addition, we present an improved *D. galeata* reference genome using Hi-C data that links the majority of scaffolds to the ten chromosomes. This contributes to the available chromosome-scale *Daphnia* assemblies (Barnard-Kubow et al., 2022; Lee et al., 2019; Wersebe et al., 2023) and is the first for the DLSC, thus allowing genome-wide investigation of the genomic landscape and the impact of hybridization and introgression.

Accurate time since admixture estimates with junctions

To understand which regions of the genome have been affected by introgression and to estimate the time since admixture on an individual basis, we have made use of the extended theory of junctions (Janzen et al., 2018; Janzen & Miró Pina, 2022). Similar approaches have used the progressive breakdown of ancestry tracts through recombination to understand genetic connectivity in the European sea bass (Duranton et al., 2019), admixture in brown trout (Leitwein et al., 2018), hybridization in swordtail fish (Schumer et al., 2018), and backcrossing in American black duck (Lavretsky et al., 2019). We used Ancestry HMM for local ancestry inference (Corbett-Detig & Nielsen, 2017), but performed time since admixture estimation using the extended theory of junctions to obtain per-individual and per-chromosome time since admixture estimates. For our data set this approach was also able to generate individual time since admixture estimates for populations with highly heterogeneous estimates across individuals and chromosomes, even when they all originated from the same time period.

The extended theory of junctions is easily adaptable to other organisms with a history of hybridization, does not require phased genomic data or a large number of markers (Janzen & Miró Pina, 2022), and can provide accurate time since admixture estimates even in the face of small population sizes. Similar to other approaches, the approach we used does require genomic reference samples used to determine ancestry informative sites and preliminary species assignment of unknown hybrids, especially when more than two species are present. Therefore, some caveats apply to the interpretation of our results. First, the recombination rate in *Daphnia* is known to vary within and between chromosomes (Wersebe et al., 2023), which is not considered in this approach. In our analysis, we explored different genome-wide *Daphnia* recombination rates, but found only minor quantitative differences (Figure S4.3), suggesting a minor effect of variation in recombination rates. Second, the limited size of the parental reference panel might have caused interference with the analysis of pure parental species due

to the incorrect identification of sites as fixed or inclusion of sites that originated from ancient introgression. We have explored different thresholds for obtaining hard ancestry calls, which should remove low-confidence sites but leave a sufficient number of sites for all samples (Table S4.5), potentially mitigating this effect.

Hybridization history of *Daphnia* populations

Hybridization dynamics are sometimes described as a pulse model, in which a single hybridization event took place between two species in the past with no further gene flow (Moran et al., 2021). While this can describe some introgression events, the history of hybridizing taxa is often much more complex, and hybridization between two or more spatially structured populations comprises multiple pulses, extends over longer time periods, and is sometimes an ongoing process (e.g., Combe et al., 2022; De La Torre et al., 2015; Vijay et al., 2016). Time-series genomic data can help us to understand pulse versus ongoing hybridization in two lake populations of the DLSC, and whether one or more parental species are present at the same time.

The additional genomic data generated here corroborate the findings of Nickel et al. (2021), with *D. galeata* × *D. longispina* hybrids in population EIC and both *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata* hybrids in population SE. With more extensive sampling and the novel junctions methods, we can now more precisely classify the number of generations since hybridization for individuals previously classified only as later-generation hybrids or backcrosses with one of the parental species. The chromosome-scale genome assembly also allows us to resolve the hybridization dynamics over time in these populations in more detail and characterize the extent and timing of admixture pulses in the genome of individuals.

While hybridization was found frequently in both populations, the lakes display different ecological conditions and histories. Eichbaumsee is an excavation lake that was built in 1976 and exhibited elevated levels of eutrophication. The extensive history of remediation efforts, e.g., the use of lanthanum-modified bentonite clay to reduce phosphate levels and hypolimnetic aeration, were ultimately unsuccessful, and high phosphate levels persist to this day (Förster et al., 2021; KLS-Gewässerschutz, 2019). In contrast, Selenter See, like many lakes in the region, was formed during the last ice age and experienced eutrophication in the middle of the 20th century caused by anthropogenic disturbances, such as fertilizers from agricultural runoff, and has since returned to its original mesotrophic state (Nixdorf et al., 2004).

The population EIC was established ~44 years prior to sampling when the lake was artificially formed. One third of the population was found to be of *D. galeata* x *D. longispina* hybrid origin based on ADMIXTURE and ancestry painting results, with the first appearance of hybrids in the history of the lake in the "core–top" group corresponding to the youngest eggs from the

sediment core. Prior to this, only a single *D. longispina* individual was found, indicating that *D. galeata* was the dominant species in the resting egg bank.

To evaluate which parts of the genome were most strongly affected by hybridization, whether any shared patterns are detectable among hybrids, and when each admixture event took place, we performed junctions analysis. In the two oldest sample groups, we found little evidence for recent admixture and the presence of the parental species *D. galeata* and *D. longispina*. However, the low time since admixture estimates for these groups and some of population SE are most likely inaccurate and instead describe parental species, based on the local ancestry genomic patterns. The "sediment–top" and "core–top" show high variation between time since admixture estimates, with each individual carrying signatures of multiple distinct admixture events and a unique evolutionary history (Figure 4.3), which is reflected by the high variation between estimates for each chromosome of an individual (Figure S4.4, Figure S4.5).

Although no genomic data for hybrid samples from older sediment layers, and thus direct evidence for hybridization, was generated, the time since admixture estimates for the "sediment–top" and "core–top" groups strongly suggest that hybrids did not appear only recently in this lake and are capable of further reproduction. The time-series genomic data therefore suggest that hybridization began when the lake was formed ~44 years prior to sampling, and is happening continuously throughout this lake population. Time since admixture estimates for two EIC samples are older than the lake (~70 years), but it is possible that some of the individuals that first colonized the lake were already of hybrid origin. The presence of both parental species in the sediment samples, presence of hybrids in the pelagic population, and previous studies (Yin et al., 2014) suggest that the majority of hybrids are produced locally, but we cannot exclude the possibility that some hybrids from other locations were able to establish in this population and contribute to the genomes we sequenced in this study.

The high level of variation also extends to local ancestry and junction patterns across the genome. Hybrids show very distinct patterns of a mosaic genomic landscape consisting of *galeata*, *longispina*, and heterospecific regions. There is little overlap between individuals in regions of high or low introgression (Figure 4.4).

All individuals in the "clones" group collected from the water column as the active pelagic population were found to be of hybrid origin and showed similar patterns in all analyses. The local ancestry across the genome and time since admixture estimates suggest that they are of very recent hybrid origin and most likely F1 due to their high heterozygosity. In general, hybrid fitness depends on the hybrid class and parental species (Arnold & Hodges, 1995), and we hypothesize that this pattern is due to hybrid vigor, where F1 hybrids are able to grow more rapidly than their parental species (Bar-Zvi et al., 2017). In *Daphnia*, this is possible due to

their clonal reproduction and has been described for both *D. galeata* × *D. longispina* (Bernatowicz et al., 2021; Griebel et al., 2015) and *D. galeata* × *D. cucullata* (Declerck & De Meester, 2003) hybrids. The *Daphnia* hybrids can outperform their parental species by being better adapted to specific temporal conditions, such as temperature and fish predation, known as the temporal hybrid superiority hypothesis (Spaak & Hoekstra, 1997).

The samples from population SE cover a shorter timescale (~26 years) and fewer individuals, but all three species were present in this lake at some point, and *D. longispina* and *D. cucullata* left traces in the genomes of some individuals. However, the only parental species identified was *D. galeata*. The individuals from population SE also show a high variation between time since admixture estimates, indicating separate admixture events, and a similar maximum estimate (~70 years) as the population EIC. We also find a mosaic genomic landscape consisting of *galeata*, *cucullata*, and heterospecific regions on single chromosomes. While we found only three *D. galeata* × *D. longispina* hybrids, they all display similar time since admixture estimates, which are on average older than those found in *D. galeata* × *D. cucullata*, but have distinct patterns across the chromosomes. This was also found for all *D. galeata* × *D. longispina* hybrids from population EIC. Overall, continuous hybridization spanning the last decades is found in both populations and species pairs.

We are able to detect general hybridization trends and patterns in these mosaic hybrid zones and hypothesize about the ecological barriers that influence them. Our time since admixture estimates are based on the assumption that all individuals have undergone one sexual reproduction every year. In reality, however, species may reproduce at different times of the year (Jankowski & Straile, 2004; Macháček et al., 2013) and survive multiple winters as asexual lineages (Griebel et al., 2016).

In naturally oligotrophic peri-Alpine lakes, the increased eutrophication in the 1970s and 1980s led to the native *D. longispina* populations being replaced by *D. galeata* and *D. galeata* × *D. longispina* hybrids (Alric et al., 2016; Brede et al., 2009; Rellstab et al., 2011). Here, we find the presence of both species, continuous hybridization, and an increased level of backcrossing with *D. longispina* in the sampled genomes of population EIC, which is a novel finding for this group. This lake is a highly anthropogenically disturbed habitat, which is known to disrupt reproductive barriers and increase hybridization rates (Crispo et al., 2011; Grabenstein & Taylor, 2018), with eutrophication and cyanobacterial blooms affecting hybridization in other species (Seehausen et al., 1997) and *Daphnia* (Alric et al., 2016; Petrusek, Seda, et al., 2008; Rellstab et al., 2011). *D. longispina* are able to perform well under high-food conditions caused by eutrophication (Spaak et al., 2012).

However, these patterns are more difficult to interpret in population SE because all three species co-occurring in the same location is rarely documented (Petrusek, Seda, et al., 2008), and random dispersal and successful species establishment may be an important factor (Holmes et al., 2016). In addition, the high dispersal rate of species may also drive hybridization (Bourret et al., 2022) and is found in *Daphnia* (Louette & De Meester, 2005).

In a previous study, we identified introgression in the DLSC and discovered that hybridization extends beyond early-generation hybrids and differs between hybrid individuals. This was difficult to interpret further without a chromosome-scale assembly to pinpoint breakpoints across a chromosome (Nickel et al., 2021), which now allows us to track ancestry states and date the time since admixture. The presence of *D. galeata* mitochondrial genomes in all *D. galeata* × *D. longispina* hybrids suggests an asymmetric maternal contribution and mostly unidirectional hybridization, but this may be caused by genetic incompatibilities or differences in the timing and frequency of sexual reproduction between species. The nuclear whole-genome patterns clearly show large regions of homospecific *D. longispina* ancestry across multiple individuals, populations, and chromosomes. This pattern of bidirectional hybridization is also found in *D. galeata* × *D. cucullata* hybrids.

Continuous hybridization over multiple generations, backcrossing with the parental species and with early hybrids, is also found in both species pairs and in both populations. The local ancestry and junction distribution across the genome show large differences between individuals, but some distinct groups with overlapping patterns; early hybrids living in the water column with largely heterozygous chromosomes, parental species with small introgressed regions stemming from not completely fixed sites or ancient introgression, later-generation hybrids (~15–30 years) with larger regions of ancestry or complete parental chromosomes, and later-generation hybrids (>50 years) with high junction density and mosaic genomic landscapes (Figure S4.6, Figure S4.7, Figure S4.8, Figure S4.9).

Few studies have looked at individual genome-wide ancestry patterns, but the evolutionary outcomes of hybridization can be highly dependent on species, environmental factors, and population history. Anthropogenic habitat disturbances can lead to contact and introgression between divergent species, such as one large region of recent introgression in *Ciona* (Fraïsse et al., 2022) or adaptive introgression in *Helicoverpa* (Valencia-Montoya et al., 2020). Genomic regions that contribute to reproductive isolation or local adaptation are less likely to introgress and prevent species from merging under these circumstances.

We find no evidence for genome stabilization, i.e., homogeneous genome composition resulting from drift or selection with fixed junction patterns across individuals (Buerkle & Rieseberg, 2008). This process can occur after relatively few generations across multiple

populations and with repeated outcomes, e.g., ~50 generations in ants (Nouhaud et al., 2022) and ~100–150 generations in swordtail fish (Langdon et al., 2022). After the initial admixture event, the long ancestry blocks are broken down by recombination in a hybrid population. Whether and over how many generations this fixation and purging of alleles occurs is modulated by ecological selection, genetic incompatibilities, and population demography (Moran et al., 2021).

Features of complex hybrid zones

In contrast to terrestrial hybrid zones, aquatic hybrid zones are less commonly studied and are often found across habitat patches with different ecological conditions, for example, in marine snails (Westram et al., 2021) and mussels (Simon et al., 2021). Freshwater hybrid zones often occur along environmental gradients in river systems (Barreto et al., 2020) or in more homogeneous lake habitats (Bittner et al., 2010). The DLSC hybrid zone combines a number of unique characteristics; the regular presence of up to two parental species without a complete collapse into a hybrid swarm, multiple species occurring in mosaic hybrid zones, and the continuous hybridization and backcrossing and multiple admixture events reflected in the genomes. We are not aware of other studies showing hybrid systems with these characteristics (Runemark et al., 2019), but our results suggest that these patterns may be present in other systems, are difficult to identify, and may be missed with no or limited genomic sampling. Even when genomic data is available, it is rarely used to track the process of hybridization in admixed individuals.

However, there are some similarities to other hybrid zones, particularly those with mosaic distributions shaped by different habitat conditions (Bell & Irian, 2019; El Ayari et al., 2019; Walsh et al., 2016). The commonly studied field cricket mosaic hybrid zone also exhibits multigenerational backcrossing and heterogeneous genomes of the sampled individuals (Harrison & Larson, 2016).

It is crucial to understand the large variation in outcomes of genomic composition in replicated hybrid zones (Gompert et al., 2017) and the mechanisms driving this variability found in models (McFarlane et al., 2022) and in nature (Mandeville et al., 2019). By using time-series genomic data of hybridization processes, we can track the genome-wide ancestry after hybridization and how selection and recombination rapidly reduce ancestry tract lengths at this early stage (Moran et al., 2021). We found the expected large variation in genome-wide ancestry among individuals with the continuous introduction of chromosomes of hybrid or parental origin.

Temporal genomic sampling has been used to study the historical movement of hybrid zones (Aguillon & Rohwer, 2022; Billerman et al., 2019; Taylor et al., 2014), but relies on previously collected and preserved samples. This temporal genomic approach will give us insights into

the rapid evolutionary response to anthropogenic change across taxa (Clark et al., 2023) and highlights the suitability of the DLSC for this purpose, where the resting egg bank can be used to directly study natural hybridizing populations without pre-existing samples.

Conclusion

In this study, we present time-series genomic data for 56 resting eggs and individuals collected from the water column to reconstruct the hybridization dynamics of two natural DLSC hybrid populations. Combined with the novel junctions method, we show that genomic patterns of hybridization and introgression can be reconstructed over several decades and applied to different hybrid systems to understand the diverse outcomes of hybridization across populations. Our results highlight the multiple successive hybridization events over several decades that shaped the diverse hybrid ancestry among individuals and across the genome. These patterns reveal the unique characteristics and complex history of a freshwater mosaic hybrid zone, not previously described in a natural hybrid population.

In the future, the genome annotation can be used to identify the biological function of introgressed genomic regions that may be under selection in local populations. This could explain the high prevalence of hybrids in these populations and point to underlying differences in ecological conditions, such as trophic and food preferences. In addition, this will also help to uncover the functional role of genes in regions of high divergence between species, and reproductive barriers in the DLSC (Nickel et al., 2021). Sampling from replicate hybrid zones with different combinations of species present will help us to determine if these patterns are universal and can be predicted.

Data accessibility

The chromosome-scale *D. galeata* genome assembly and annotation are deposited in Zenodo (doi:10.5281/zenodo.8069314) and will be stored in NCBI GenBank. All resequencing data are deposited at the European Nucleotide Archive under project accession number PRJEB60365. Previously generated sequencing data used here is available under PRJEB40405. Individual accession numbers are provided in Table S4.2.

Acknowledgments

We thank the Institut für Hygiene und Umwelt for providing us access to Eichbaumsee, and Fynn Eilers, Jasper Kremer, and Jennifer Drechsler for their assistance in the molecular laboratory and with sampling.

Chapter 4 References

Aguillon, S. M., & Rohwer, V. G. (2022). Revisiting a classic hybrid zone: Movement of the northern flicker hybrid zone in contemporary times. *Evolution*, *76*(5), 1082–1090. https://doi.org/10.1111/evo.14474

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Alric, B., Möst, M., Domaizon, I., Pignol, C., Spaak, P., & Perga, M.-E. (2016). Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology*, *29*(4), 720–735. https://doi.org/10.1111/jeb.12820

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnold, M. L., & Hodges, S. A. (1995). Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution*, *10*(2), 67–71. https://doi.org/10.1016/S0169-5347(00)88979-X

Barnard-Kubow, K. B., Becker, D., Murray, C. S., Porter, R., Gutierrez, G., Erickson, P., Nunez, J. C. B., Voss, E., Suryamohan, K., Ratan, A., Beckerman, A., & Bergland, A. O. (2022). Genetic Variation in Reproductive Investment Across an Ephemerality Gradient in *Daphnia pulex*. *Molecular Biology and Evolution*, *39*(6), msac121. https://doi.org/10.1093/molbev/msac121

Barreto, S. B., Knowles, L. L., Affonso, P. R. A. D. M., & Batalha-Filho, H. (2020). Riverscape properties contribute to the origin and structure of a hybrid zone in a Neotropical freshwater fish. *Journal of Evolutionary Biology*, *33*(11), 1530–1542. https://doi.org/10.1111/jeb.13689

Barth, J. M. I., Gubili, C., Matschiner, M., Tørresen, O. K., Watanabe, S., Egger, B., Han, Y.-S., Feunteun, E., Sommaruga, R., Jehle, R., & Schabetsberger, R. (2020). Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15099-x

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid Zones and the Evolutionary Process* (pp. 13–45). Oxford University Press.

Bar-Zvi, D., Lupo, O., Levy, A. A., & Barkai, N. (2017). Hybrid vigor: The best of both parents, or a genomic clash? *Current Opinion in Systems Biology*, *6*, 22–27. https://doi.org/10.1016/j.coisb.2017.08.004

Beaton, M. J., & Hebert, P. D. N. (1994). Variation in chromosome numbers of *Daphnia* (Crustacea, Cladocera). *Hereditas*, *120*(3), 275–279. https://doi.org/10.1111/j.1601-5223.1994.00275.x

Bell, R. C., & Irian, C. G. (2019). Phenotypic and genetic divergence in reed frogs across a mosaic hybrid zone on São Tomé Island. *Biological Journal of the Linnean Society*, *128*(3), 672–680. https://doi.org/10.1093/biolinnean/blz131

Beninde, J., Feldmeier, S., Veith, M., & Hochkirch, A. (2018). Admixture of hybrid swarms of native and introduced lizards in cities is determined by the cityscape structure and invasion history. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883), 20180143. https://doi.org/10.1098/rspb.2018.0143

Bernatowicz, P., Dawidowicz, P., & Pijanowska, J. (2021). Phenotypic plasticity and developmental noise in hybrid and parental clones of *Daphnia longispina* complex. *Aquatic Ecology*, *55*(4), 1179–1188. https://doi.org/10.1007/s10452-021-09898-7

Billerman, S. M., Cicero, C., Bowie, R. C. K., & Carling, M. D. (2019). Phenotypic and genetic introgression across a moving woodpecker hybrid zone. *Molecular Ecology*, *28*(7), 1692–1708. https://doi.org/10.1111/mec.15043

Bittner, D., Excoffier, L., & Largiadèr, C. R. (2010). Patterns of morphological changes and hybridization between sympatric whitefish morphs (*Coregonus* spp.) in a Swiss lake: A role for eutrophication? *Molecular Ecology*, *19*(10), 2152–2167. https://doi.org/10.1111/j.1365-294X.2010.04623.x

Blackwell, T., Ford, A. G. P., Ciezarek, A. G., Bradbeer, S. J., Juarez, C. A. G., Smith, A. M., Ngatunga, B. P., Shechonge, A., Tamatamah, R., Etherington, G., Haerty, W., Palma, F. D., Turner, G. F., & Genner, M. J. (2021). Newly discovered cichlid fish biodiversity threatened by hybridization with non-native species. *Molecular Ecology*, *30*(4), 895–911. https://doi.org/10.1111/mec.15638

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bourret, S. L., Kovach, R. P., Cline, T. J., Strait, J. T., & Muhlfeld, C. C. (2022). High dispersal rates in hybrids drive expansion of maladaptive hybridization. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1986), 20221813. https://doi.org/10.1098/rspb.2022.1813

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species.

Proceedings of the National Academy of Sciences, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Brendonck, L., & De Meester, L. (2003). Egg banks in freshwater zooplankton: Evolutionary and ecological archives in the sediment. *Hydrobiologia*, *491*(1), 65–84. https://doi.org/10.1023/A:1024454905119

Buerkle, C. A., & Rieseberg, L. H. (2008). The rate of genome stabilization in homoploid hybrid species. *Evolution*, *62*(2), 266–275. https://doi.org/10.1111/j.1558-5646.2007.00267.x

Cáceres, C. E. (1998). Interspecific Variation in the Abundance, Production, and Emergence of *Daphnia* Diapausing Eggs. *Ecology*, *79*(5), 1699–1710. https://doi.org/10.1890/0012-9658(1998)079[1699:IVITAP]2.0.CO;2

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Canestrelli, D., Bisconti, R., Chiocchio, A., Maiorano, L., Zampiglia, M., & Nascetti, G. (2017). Climate change promotes hybridisation between deeply divergent species. *PeerJ*, *5*, e3072. https://doi.org/10.7717/peerj.3072

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chaturvedi, S., Lucas, L. K., Buerkle, C. A., Fordyce, J. A., Forister, M. L., Nice, C. C., & Gompert, Z. (2020). Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications*, *11*(1), 2179. https://doi.org/10.1038/s41467-020-15641-x

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chhatre, V. E., Evans, L. M., DiFazio, S. P., & Keller, S. R. (2018). Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Molecular Ecology*, *27*(23), 4820–4838. https://doi.org/10.1111/mec.14820

Chin, T. A., & Cristescu, M. E. (2021). Speciation in *Daphnia*. *Molecular Ecology*, *30*(6), 1398–1418. https://doi.org/10.1111/mec.15824

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., Sawkins, A., Snead, A. A., Whalen, J. C., & Pinsky, M. L. (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13789

Combe, F. J., Jaster, L., Ricketts, A., Haukos, D., & Hope, A. G. (2022). Population genomics of free-ranging Great Plains white-tailed and mule deer reflects a long history of interspecific hybridization. *Evolutionary Applications*, *15*(1), 111–131. https://doi.org/10.1111/eva.13330

Corbett-Detig, R., & Nielsen, R. (2017). A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLoS Genetics*, *13*(1), e1006529. https://doi.org/10.1371/journal.pgen.1006529

Cornetti, L., Fields, P. D., Van Damme, K., & Ebert, D. (2019). A fossil-calibrated phylogenomic analysis of *Daphnia* and the Daphniidae. *Molecular Phylogenetics and Evolution*, *137*, 250–262. https://doi.org/10.1016/j.ympev.2019.05.018

Coulter, A. A., Brey, M. K., Lamer, J. T., Whitledge, G. W., & Garvey, J. E. (2020). Early generation hybrids may drive range expansion of two invasive fishes. *Freshwater Biology*, *65*(4), 716–730. https://doi.org/10.1111/fwb.13461

Crispo, E., Moore, J.-S., Lee-Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals. *BioEssays*, *33*(7), 508–518. https://doi.org/10.1002/bies.201000154

Cristescu, M. E. A., Colbourne, J. K., Radivojac, J., & Lynch, M. (2006). A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex*: On the prospect of crustacean genomics. *Genomics*, *88*(4), 415–430. https://doi.org/10.1016/j.ygeno.2006.03.007

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

De La Torre, A., Ingvarsson, P. K., & Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, *115*(2), 153–164. https://doi.org/10.1038/hdy.2015.19

Declerck, S., & De Meester, L. (2003). Impact of fish predation on coexisting *Daphnia* taxa: A partial test of the temporal hybrid superiority hypothesis. *Hydrobiologia*, *500*(1), 83–94. https://doi.org/10.1023/A:1024656602248

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95. https://doi.org/10.1126/science.aal3327

Dukić, M., Berner, D., Roesti, M., Haag, C. R., & Ebert, D. (2016). A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC Genetics*, *17*(1), 137. https://doi.org/10.1186/s12863-016-0445-7

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems*, *3*(1), 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

Duranton, M., Bonhomme, F., & Gagnaire, P.-A. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications*, *12*(9), 1743–1756. https://doi.org/10.1111/eva.12829

Ebert, D. (2005). *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*. National Library of Medicine (US), National Center for Biotechnology. https://www.ncbi.nlm.nih.gov/books/NBK2036/

Ebert, D. (2022). *Daphnia* as a versatile model system in ecology and evolution. *EvoDevo*, *13*(1), 16. https://doi.org/10.1186/s13227-022-00199-0

El Ayari, T., Trigui El Menif, N., Hamer, B., Cahill, A. E., & Bierne, N. (2019). The hidden side of a major marine biogeographic boundary: A wide mosaic hybrid zone at the Atlantic–Mediterranean divide reveals the complex interaction between natural and genetic barriers in mussels. *Heredity*, *122*(6), 770–784. https://doi.org/10.1038/s41437-018-0174-y

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Förster, W., Scholten, J. C., Schubert, M., Knoeller, K., Classen, N., Lechelt, M., Richard, J.-H., Rohweder, U., Zunker, I., & Wanner, S. C. (2021). Phosphorous Supply to a Eutrophic Artificial Lake: Sedimentary versus Groundwater Sources. *Water*, *13*(4), 563. https://doi.org/10.3390/w13040563

Fraïsse, C., Le Moan, A., Roux, C., Dubois, G., Daguin-Thiebaut, C., Gagnaire, P.-A., Viard, F., & Bierne, N. (2022). Introgression between highly divergent sea squirt genomes: An adaptive breakthrough? *Peer Community Journal*, *2*, e54. https://doi.org/10.24072/pcjournal.172

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368. https://doi.org/10.1111/ele.12237

Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of Population Genomic Data from Hybrid Zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207–229. https://doi.org/10.1146/annurev-ecolsys-110316-022652

Gompert, Z., Springer, A., Brady, M., Chaturvedi, S., & Lucas, L. K. (2021). Genomic time-series data show that gene flow maintains high genetic diversity despite substantial genetic drift in a butterfly species. *Molecular Ecology*, *30*(20), 4991–5008. https://doi.org/10.1111/mec.16111

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking Barriers: Causes, Consequences, and Experimental Utility of Human-Mediated Hybridization. *Trends in Ecology & Evolution*, *33*(3), 198–212. https://doi.org/10.1016/j.tree.2017.12.008

Griebel, J., Gießler, S., Poxleitner, M., Faria, A. N., Yin, M., & Wolinska, J. (2015). Extreme Environments Facilitate Hybrid Superiority – The Story of a Successful *Daphnia galeata* × *longispina* Hybrid Clone. *PLoS One*, *10*(10), e0140275. https://doi.org/10.1371/journal.pone.0140275

Griebel, J., Gießler, S., Yin, M., & Wolinska, J. (2016). Parental and hybrid *Daphnia* from the *D. longispina* complex: Long-term dynamics in genetic structure and significance of overwintering modes. *Journal of Evolutionary Biology*, *29*(4), 810–823. https://doi.org/10.1111/jeb.12828

Holmes, C. J., Pantel, J. H., Schulz, K. L., & Cáceres, C. E. (2016). Initial genetic diversity enhances population establishment and alters genetic structuring of a newly established *Daphnia* metapopulation. *Molecular Ecology*, *25*(14), 3299–3308. https://doi.org/10.1111/mec.13672

Huylmans, A. K., López Ezquerra, A., Parsch, J., & Cordellier, M. (2016). De Novo Transcriptome Assembly and Sex-Biased Gene Expression in the Cyclical Parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution*, *8*(10), 3120–3139. https://doi.org/10.1093/gbe/evw221

Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, *30*(17), 4245–4258. https://doi.org/10.1111/mec.16056

Jankowski, T., & Straile, D. (2004). Allochronic differentiation among *Daphnia* species, hybrids and backcrosses: The importance of sexual reproduction for population dynamics and genetic architecture. *Journal of Evolutionary Biology*, *17*(2), 312–321.

Janzen, T. (2021). *Junctions: The Breakdown of Genomic Ancestry Blocks in Hybrid Lineages.* (R package version 2.0.3). https://github.com/thijsjanzen/junctions

Janzen, T., & Miró Pina, V. (2022). Estimating the time since admixture from phased and unphased molecular data. *Molecular Ecology Resources*, *22*(3), 908–926. https://doi.org/10.1111/1755-0998.13519

Janzen, T., Nolte, A. W., & Traulsen, A. (2018). The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. *Evolution*, *72*(4), 735–750. https://doi.org/10.1111/evo.13436

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., & Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, *19*, 189. https://doi.org/10.1186/s12859-018-2203-5

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, *44*(9), e89. https://doi.org/10.1093/nar/gkw092

Keller, B., Wolinska, J., Manca, M., & Spaak, P. (2008). Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2943–2952. https://doi.org/10.1098/rstb.2008.0044

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

KLS-Gewässerschutz. (2019). *Hamburgs Badegwässer – Badegewässerprofil Eichbaumsee*. Freie und Hansestadt Hamburg, Behörde für Umwelt und Energie, Amt für Wasser, Abwasser und Geologie. https://www.hamburg.de/contentblob/3348608/311e578d31e9c4dd2f6e575e8e359537/data/d-profil-eichbaumsee.pdf

Kong, S., & Kubatko, L. S. (2021). Comparative Performance of Popular Methods for Hybrid Detection using Genomic Data. *Systematic Biology*, *70*(5), 891–907. https://doi.org/10.1093/sysbio/syaa092

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Langdon, Q. K., Powell, D. L., Kim, B., Banerjee, S. M., Payne, C., Dodge, T. O., Moran, B., Fascinetto-Zago, P., & Schumer, M. (2022). Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLoS Genetics*, *18*(1), e1009914. https://doi.org/10.1371/journal.pgen.1009914

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, *9*(6), 3470–3490. https://doi.org/10.1002/ece3.4981

Le Moan, A., Roby, C., Fraïsse, C., Daguin-Thiébaut, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, *30*(24), 6718–6732. https://doi.org/10.1111/mec.16189

Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, *210*, 69–84. https://doi.org/10.1016/j.aquatox.2019.02.009

Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, *27*(17), 3466–3483. https://doi.org/10.1111/mec.14816

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*. https://doi.org/10.48550/arXiv.1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Louette, G., & De Meester, L. (2005). High Dispersal Capacity of Cladoceran Zooplankton in Newly Founded Communities. *Ecology*, *86*(2), 353–359. https://doi.org/10.1890/04-0403

Mandeville, E. G., Walters, A. W., Nordberg, B. J., Higgins, K. H., Burckhardt, J. C., & Wagner, C. E. (2019). Variable hybridization outcomes in trout are predicted by historical fish stocking and

environmental context. *Molecular Ecology*, *28*(16), 3738–3755. https://doi.org/10.1111/mec.15175

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

McFarlane, S. E., Jahner, J. P., Lindtke, D., Buerkle, C. A., & Mandeville, E. G. (2022). Selection leads to remarkable variability in the outcomes of hybridization across replicate hybrid zones. *BioRxiv*. https://doi.org/10.1101/2022.09.23.509250

McFarlane, S. E., Senn, H. V., Smith, S. L., & Pemberton, J. M. (2021). Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, *30*(9), 2104–2115. https://doi.org/10.1111/mec.15862

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Meng, G., Li, Y., Yang, C., & Liu, S. (2019). MitoZ: A toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Research*, *47*(11), e63–e63. https://doi.org/10.1093/nar/gkz173

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *ELife*, *10*, e69016. https://doi.org/10.7554/eLife.69016

Natola, L., Seneviratne, S. S., & Irwin, D. (2022). Population genomics of an emergent tri-species hybrid zone. *Molecular Ecology*, *31*(20), 5356–5367. https://doi.org/10.1111/mec.16650

Nickel, J., & Cordellier, M. (2022). Cost-saving population genomic investigation of *Daphnia longispina* complex resting eggs using whole-genome amplification and pre-sequencing screening. *Ecology and Evolution*, *12*(12), e9682. https://doi.org/10.1002/ece3.9682

Nickel, J., Schell, T., Holtzem, T., Thielsch, A., Dennis, S. R., Schlick-Steiner, B. C., Steiner, F. M., Möst, M., Pfenninger, M., Schwenk, K., & Cordellier, M. (2021). Hybridization Dynamics and Extensive Introgression in the *Daphnia longispina* Species Complex: New Insights from a High-Quality *Daphnia galeata* Reference Genome. *Genome Biology and Evolution*, *13*(12), evab267. https://doi.org/10.1093/gbe/evab267

Nixdorf, B., Hemm, M., Hoffmann, A., & Richter, P. (2004). *Dokumentation von Zustand und Entwicklung der wichtigsten Seen Deutschlands*. Umweltbundesamt. https://www.umweltbundesamt.de/publikationen/dokumentation-von-zustand-entwicklung-wichtigsten

Nouhaud, P., Martin, S. H., Portinha, B., Sousa, V. C., & Kulmuni, J. (2022). Rapid and predictable genome evolution across three hybrid ant populations. *PLoS Biology*, *20*(12), e3001914. https://doi.org/10.1371/journal.pbio.3001914

O'Grady, C. J., Dhandapani, V., Colbourne, J. K., & Frisch, D. (2022). Refining the evolutionary time machine: An assessment of whole genome amplification using single historical *Daphnia* eggs. *Molecular Ecology Resources*, *22*(3), 946–961. https://doi.org/10.1111/1755-0998.13524

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J. K., Pfrender, M. E., & Weider, L. J. (2013). The evolutionary time machine: Using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution*, *28*(5), 274–282. https://doi.org/10.1016/j.tree.2013.01.009

Ottenburghs, J. (2019). Multispecies hybridization in birds. *Avian Research*, *10*(1), 20. https://doi.org/10.1186/s40657-019-0159-4

Petrusek, A., Hobæk, A., Nilssen, J. P., Skage, M., Černý, M., Brede, N., & Schwenk, K. (2008). A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta*, *37*(5), 507–519. https://doi.org/10.1111/j.1463-6409.2008.00336.x

Petrusek, A., Seda, J., Macháček, J., Ruthová, Š., & Šmilauer, P. (2008). *Daphnia* hybridization along ecological gradients in pelagic environments: The potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2931–2941. https://doi.org/10.1098/rstb.2008.0026

Pietrzak, B., & Ślusarczyk, M. (2006). The fate of the ephippia – *Daphnia* dispersal in time and space. *Polish Journal of Ecology*, *54*(4), 709–714.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., Baalsrud, H. T., Brieuc, M. S. O., Kjesbu, O. S., Godiksen, J. A., Barth, J. M. I., Matschiner, M., Stenseth, N. C., Jakobsen, K. S., Jentoft, S., & Star, B. (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences*, *118*(15), e2025453118. https://doi.org/10.1073/pnas.2025453118

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. https://doi.org/10.1101/201178

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rellstab, C., Keller, B., Girardclos, S., Anselmetti, F. S., & Spaak, P. (2011). Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. *Limnology and Oceanography*, *56*(1), 292–302. https://doi.org/10.4319/lo.2011.56.1.0292

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., Elgvin, T. O., & Sætre, G.-P. (2018). Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution*, *2*(3), 549–556. https://doi.org/10.1038/s41559-017-0437-7

Runemark, A., Vallejo-Marin, M., & Meier, J. I. (2019). Eukaryote hybrid genomes. *PLoS Genetics*, *15*(11), e1008404. https://doi.org/10.1371/journal.pgen.1008404

Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences*, *115*(10), E2284–E2291. https://doi.org/10.1073/pnas.1714950115

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, *17*(1), 30–44. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Seehausen, O., Van Alphen, J. J. M., & Witte, F. (1997). Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection. *Science*, *277*(5333), 1808–1811. https://doi.org/10.1126/science.277.5333.1808

Simon, A., Fraïsse, C., El Ayari, T., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021). How do species barriers decay? Concordance and local introgression in mosaic hybrid zones of mussels. *Journal of Evolutionary Biology*, *34*(1), 208–223. https://doi.org/10.1111/jeb.13709

Spaak, P., Denk, A., Boersma, M., & Weider, L. J. (2004). Spatial and temporal patterns of sexual reproduction in a hybrid *Daphnia* species complex. *Journal of Plankton Research*, *26*(6), 625–635. https://doi.org/10.1093/plankt/fbh064

Spaak, P., Fox, J., & Hairston, N. G. (2012). Modes and mechanisms of a *Daphnia* invasion. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2936–2944. https://doi.org/10.1098/rspb.2012.0280

Spaak, P., & Hoekstra, J. R. (1997). Fish predation on a *Daphnia* hybrid species complex: A factor explaining species coexistence? *Limnology and Oceanography*, *42*(4), 753–762. https://doi.org/10.4319/lo.1997.42.4.0753

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, *25*(11), 2427–2442. https://doi.org/10.1111/mec.13539

Taylor, S. A., Curry, R. L., White, T. A., Ferretti, V., & Lovette, I. (2014). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, *68*(11), 3066–3081. https://doi.org/10.1111/evo.12510

Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177. https://doi.org/10.1038/s41559-018-0777-y

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., Gordon, K. H. J., Specht, A., Paula-Moraes, S. V., Rane, R., Walsh, T. K., & Jiggins, C. D. (2020). Adaptive Introgression across Semipermeable Species Boundaries between Local *Helicoverpa zea* and Invasive *Helicoverpa armigera* Moths. *Molecular Biology and Evolution*, *37*(9), 2568–2583. https://doi.org/10.1093/molbev/msaa108

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*(1), 13195. https://doi.org/10.1038/ncomms13195

Walsh, J., Rowe, R. J., Olsen, B. J., Shriver, W. G., & Kovach, A. I. (2016). Genotype-environment associations support a mosaic hybrid zone between two tidal marsh birds. *Ecology and Evolution*, *6*(1), 279–294. https://doi.org/10.1002/ece3.1864

Wersebe, M. J., Sherman, R. E., Jeyasingh, P. D., & Weider, L. J. (2023). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology*, *32*(6), 1478–1496. https://doi.org/10.1111/mec.16383

Westram, A. M., Faria, R., Johannesson, K., & Butlin, R. (2021). Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, *30*(15), 3797–3814. https://doi.org/10.1111/mec.15861

Yin, M., Gießler, S., Griebel, J., & Wolinska, J. (2014). Hybridizing *Daphnia* communities from ten neighbouring lakes: Spatio-temporal dynamics, local processes, gene flow and invasiveness. *BMC Evolutionary Biology*, *14*(1), 80. https://doi.org/10.1186/1471-2148-14-80

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Chapter 5 – General discussion

"We used to make fun of Edgar Anderson by saying that he was finding hybrids under every bush. Then we realized that even the bushes were hybrids."

(Abbott et al., 2013)

The focus of this thesis was to understand (i) hybridization dynamics in the *Daphnia longispina* species complex (DLSC) on a genomic level over time and (ii) the outcome of recent hybridization in mosaic hybrid zones. I highlight the implications of our findings for complex hybrid zones and discuss future perspectives.

In **Chapter 2**, we developed a method that facilitates the genomic investigation of the *Daphnia* resting egg bank by testing multiple protocols for resting egg isolation and whole genome amplification (WGA). This allows for broader sampling compared to previous methods consisting of hatching eggs from the resting egg bank and establishing clonal lines in the laboratory. Further, it enables a more accurate representation of hybrids with lower hatching success. We have established a contamination screening using simple PCR markers that reduces costs by testing samples before library preparation and sequencing steps and have shown that the resulting whole-genome data is suitable for unbiased population genomic analysis.

We applied this method directly to sequence 49 whole DLSC genomes from four natural populations in northern Germany in **Chapter 3**. To analyze this data, we produced the first high-quality genome assembly and annotation for a member of the DLSC, *Daphnia galeata*. Thus, we were able to move from the limited genetic markers previously used to the genome-wide investigation of divergence patterns between the three species. We uncovered a complex history of hybridization and introgression, reflecting multiple generations of hybridization and backcrossing in two populations, and evidence for cytonuclear discordance.

The *D. galeata* genome was further refined to a chromosome scale with Hi-C data in **Chapter 4**. To understand the introgression patterns uncovered in the two populations in **Chapter 3**, more extensive time-series genomic data was generated from single resting eggs collected from different time periods and individuals living in the water column. Using the novel "junctions" method to estimate when each hybridization event took place, we found multiple successive hybridization events over several decades and diverse genome-wide hybrid ancestry among individuals in a unique mosaic freshwater hybrid zone.

Understanding Hybrid Zones

The historical distribution of taxa in hybrid zone research has often been uneven, with biologists tending to focus on more “charismatic” taxa that are easier to identify and sample, such as certain insects, birds, or plants. In the recent past, this collection of species has been expanded to include more diverse systems, facilitated by the development of new genomic tools to study the evolutionary processes that shape hybrid genomes. The heterogeneous genomes in hybrid zones are influenced by variation in selection, recombination rate, and gene flow (Harrison & Larson, 2016). Recent hybrid zones provide insight into these early stages and species barriers. As discussed in the general introduction, the potential evolutionary outcomes of hybridization are diverse and are influenced by various biological processes that vary between hybridizing species and about which we often know very little.

While hybridization is often discussed, modeled, and analyzed as a pulse model, in which a single hybridization event took place between two species in the past with no further gene flow (B. M. Moran et al., 2021), increasingly complex models with recurring pulses or ongoing hybridization are being considered. In some ecosystems, recurrent admixture has been documented over long periods of time, e.g., in woodrats (Jahner et al., 2021) and deer (McFarlane et al., 2020). We were able to detect multiple and successive admixture pulses in the genome of DLSC individuals using the extended theory of junctions (Janzen et al., 2018; Janzen & Miró Pina, 2022).

Another aspect that researchers are beginning to explore is more complex hybrid zone dynamics, e.g., the presence of more than two species (Grummer et al., 2021; Satokangas et al., 2023) or multiple contact zones (Johannesson et al., 2020; Riquet et al., 2019). This allows us to compare genomic barriers to gene flow between species and populations with different histories.

Natola et al. (2022) found that hybridization was more extensive in tri-species sapsucker hybrid zones than in di-species hybrid zones, and can include hybrids with ancestry from more than two species, but boundaries between the sapsucker species were largely maintained. Chhatre et al. (2018) identified recurrent hybridization in range-edge tri-species *Populus* hybrid zones and found evidence for adaptive introgression. Most likely, multi-species hybrid zones are more common in sympatric taxa with low reproductive isolation (Ottenburghs, 2019).

Hybrid zones are not static, and spatio-temporal changes of hybrid zones have been observed in many ecosystems (Aguillon & Rohwer, 2022; Ryan et al., 2018; Smith et al., 2013). Traces in the genome and direct sampling over time show that hybrid zone movement is pervasive and shapes past and present hybrid zones (Taylor et al., 2015; Wielstra, 2019). Furthermore,

hybridization can expand rapidly following biological invasions and form ancestry clines across broad climatic ranges (Calfee et al., 2020).

Hybridization frequency, meaning either in how many populations hybrids are found or how many individuals in the population have a hybrid origin, is highly variable between taxa. For example, Zbinden et al. (2023) found hybridization in 56% of the sampled fish populations, with hybrid individuals ranging from 0−4.4% depending on the freshwater fish family. Hybrid frequency within populations was 15−30% in *Helicoverpa* moths (Cordeiro et al., 2020) and up to 22.6% in tropical eel species (Barth et al., 2020). Our findings in *Daphnia* indicate that 33.3% of individuals in the Eichbaumsee (EIC) population and 64.3% of individuals in the Selenter See (SE) population have hybrid ancestry. This seems relatively high compared to other taxa, but can be common during specific periods in *Daphnia* populations (Brede et al., 2009; Keller et al., 2008).

Importantly, the majority of our knowledge of hybrid zones stems from studies of terrestrial hybrid zones, which are often easier to study and are characterized by sharp clines of trait variation in clearly defined geographic regions. In aquatic hybrid zones, the different ecological factors that influence these ecosystems often lead to a mosaic distribution of hybrids, e.g., in marine environments through human-mediated transport (Le Moan et al., 2021; Simon et al., 2020) or ongoing hybridization in anadromous fish (Rougemont et al., 2022). Other factors to consider are the high reproductive rate and dispersal ability of aquatic organisms (Viard et al., 2020). In freshwater hybrid zones, most studies focus on river and stream systems, with only few examples for the more homogeneous lake and pond habitats, such as whitefish (De-Kayne et al., 2022), cichlids (Lewanski et al., 2022), and *Daphnia* (**Chapter 3** and **4**).

All these aspects add different dimensions of complexity and are important to consider when identifying hybrids and analyzing genomic ancestries. I propose that these complex hybrid zones with continuous hybridization, as found in the DLSC, are not a rare phenomenon, but are more difficult to identify and therefore often overlooked. This challenges our assumptions about how genome-wide differences in ancestry arise in the first generations after hybridization (Veller et al., 2023), the effects of hybrid incompatibilities and selection, and the evolutionary outcomes of hybridization.

Temporal genomic data and the extended theory of junctions give us insight into how hybridization has played out in the past. In addition, current genomic patterns can also give us clues about contemporary and ancient hybridization. Replicated hybridization events between the same species are predicted to be modulated by the same genetic interactions and selective forces and therefore lead to potentially similar and predictable outcomes at the genomic level (B. M. Moran et al., 2021). That these genomic outcomes are predictable in nature has been

shown in swordtail fish (Langdon et al., 2022) and *Lycaeides* butterflies, where ancestry patterns in ancient hybrids are predictable from contemporary hybrids (S. Chaturvedi et al., 2020). Nouhaud et al. (2022) found that this can occur after only ~50 generations in wood ants and is driven by recombination rate variation and genetic load. The DLSC is a promising system to explore this further, as we find extensive recent hybridization (**Chapter 3** and **4**) with genomic data from parental species, early hybrids living in the water column, and later-generation hybrids (>50 years), as well as ancient hybridization (Beninde, 2021).

The *Daphnia longispina* Species Complex as a Hybrid System

Methodological advancements

Daphnia is often described as a genomic model organism (Altshuler et al., 2011), and looking at the eight different *Daphnia* species for which at least one genome is available on NCBI (last accessed May 2023), it certainly seems that way. However, given these genomic resources, it is surprising how comparatively few studies have used them to understand the evolution of natural populations. Some of the key methods needed were missing or insufficient, but the new insights gained about *Daphnia* and improvements in genomic methods in this thesis lay the foundation for the DLSC as a genomic model organism, especially in the emerging field of speciation genomics (Seehausen et al., 2014).

While effective methods have been developed to sequence DNA from single adult *D. pulicaria* (Lack et al., 2018) and different small-bodied zooplankton species, including the DLSC (Beninde et al., 2020), this remains challenging for single resting eggs with much fewer cells. We initially identified the issue of failed amplification of resting eggs in **Chapter 2**, but we were unable to identify one or more specific organisms as major contaminants, and most MiSeq reads could not be mapped to any other available genome. These samples could not be reliably identified in our pre-sequencing contamination screening, which uses bacterial and *Daphnia* markers. This was also the case for the failed amplification of resting eggs with higher sequencing coverage in **Chapter 3** and **4**, and two out of the three failed eggs in Lack et al. (2018). I propose that this is due to “junk” DNA produced by the phi29 polymerase during MDA-WGA when the DNA is degraded or the amount is too small (Nelson, 2014), and we have not found a way to reliably detect this issue in samples prior to sequencing. One way to address this would be to use real-time PCR to directly quantify the amount of *Daphnia* DNA in the samples or to perform ultra-low coverage whole-genome sequencing and only proceed to high coverage sequencing with successful samples.

We used the successfully established WGA protocol in **Chapter 3** and **4** for resting eggs collected from sediments up to ~44 years old. A total of 112 WGA samples were sequenced after pre-sequencing contamination checks and 66 samples from four populations were

identified as *Daphnia* reads (59% success rate). This is lower than the success rate in **Chapter 2**, where for 88% of the sequenced samples *Daphnia* DNA was successfully amplified. However, only recent EIC resting eggs were used in this study, and the egg quality, which is linked to environmental conditions in the lake and sediment age, seems to have a large effect on the success rate. This method is still labor-intensive as many ephippia are empty or the WGA product is classified as contaminated, for example, over 1,000 resting eggs were collected from sediment for a total of 36 EIC genomes, but it was still possible to sequence eggs from the entire history of the lake.

O'Grady et al. (2022) generally achieved higher success rates for WGA of *D. magna* eggs from the laboratory (86%) and *D. pulicaria* from Arctic lake sediments up to ~180 years old (78%). This could be related to additional washing steps of the isolated resting eggs or different WGA kits used. In addition, the generally larger size of these species and thus larger resting eggs with a higher number of cells compared to the DLSC may play a role in the higher success rate (Möst M, personal communication). The WGA method and pre-sequencing contamination screening with adapted primers could also be applied to resting stages of other organisms with low amounts of DNA, such as other Cladocera taxa, rotifers, or diatoms. This would provide a more complete understanding of populations in freshwater ecosystems.

Overall, this method allows for the genomic investigation of the resting egg bank and advances the field of temporal genomics (Clark et al., 2023), which typically requires previously collected and preserved samples to directly examine the evolutionary response to anthropogenic changes over the past 200 years. These biological archives are directly available in this group, which is highly affected by anthropogenic changes in freshwater ecosystems (Frisch et al., 2014; Orsini et al., 2012; Rellstab et al., 2011). In addition, genomic data generated with this method for population Dobersdorfer See (DOB) was also used to estimate the effective population size of *D. galeata* (Pfenninger et al., 2022).

For *Daphnia,* it is also crucial to sample the active pelagic populations in lakes, as the resting egg bank population, which represents successful sexual reproduction and therefore likely contribution to the next generations, may be different due to differences in sexual reproduction between taxa, especially for hybrids (Jankowski & Straile, 2003; Vaníčková et al., 2010). Thus, hybrid individuals in the pelagic populations represent genotypes with generally high fitness that were able to reproduce and hatch.

The analysis of whole-genome sequences requires a genome assembly for a closely related species. So far, this was missing for the DLSC, and we present a *D. galeata* reference genome using long- and short-read sequence data for highly contiguous and accurate scaffolds that can be used for all species. Then, we improved the genome using Hi-C data to achieve

chromosome-scale scaffolds to identify recombination breakpoints along each chromosome and better understand genome-wide hybridization dynamics. This allowed us to link 96.6% of the total assembly length to the ten chromosomes and further improve the N50 value, similar to those of other published chromosome-scale *Daphnia* assemblies (Barnard-Kubow et al., 2022; Lee et al., 2019; Wersebe et al., 2023). The small genome size of *D. galeata* (~133 Mb) makes low-coverage whole-genome sequencing of multiple populations relatively accessible instead of reduced representation approaches.

Finally, we present a development of the extended theory of junctions, which is based on how recombination breaks up genomic blocks of contiguous ancestry over time, to understand the timing of hybridization dynamics (Janzen et al., 2018; Janzen & Miró Pina, 2022). Similar approaches have been used to understand genetic connectivity in European sea bass (Duranton et al., 2019), admixture in brown trout (Leitwein et al., 2018), and backcrossing in American black duck (Lavretsky et al., 2019). This method was well suited for our data set, and we were able to obtain per-individual and per-chromosome time since admixture estimates for populations with highly heterogeneous estimates.

However, this method is sensitive to recombination rate assumptions, and we tested this using different *Daphnia* species. We chose to use the recombination rate for *D. pulicaria* because it is not available for *D. galeata* or a closely related species and is known to vary within and between chromosomes (Wersebe et al., 2023). In addition, the limited size of the parental reference panel used to determine ancestry informative sites may not reflect the species accurately and incorrectly identified sites as fixed or included sites affected by ancient introgression. Our results support that this approach produces stable time since admixture estimates and can be used to understand past populations even in the absence of temporal genomic data. This was also found in the reanalysis of empirical data sets of *Saccharomyces cerevisiae*, swordtail fish, and *Populus* trees with junctions, which were consistent with previous time since admixture estimates (Janzen & Miró Pina, 2022).

This approach is easily adaptable for other hybrid systems with ancestry informative sites and does not require a large number of markers (Janzen & Miró Pina, 2022). This has the potential to identify the large variability in hybrid populations and the extent and timing of hybridization pulses, which is often overlooked when populations are analyzed together. One relevant application is conservation decisions for hybrids, which often require genomic data and the number of generations since hybridization (VonHoldt et al., 2018), but are limited to determining this only for individuals in a few specific species, such as European wildcats (Mattucci et al., 2019) and wolves (Galaverni et al., 2017).

In conclusion, this thesis and its contributions represent the transition from low-resolution genetic markers to whole genomes from different time periods and open up new possibilities to study evolutionary changes and hybridization in the DLSC.

An overview of *Daphnia* hybridization dynamics

Our comprehensive data set consisted of 72 whole genomes from four populations with different ecological conditions and 12 whole genomes as reference species, including *D. galeata*, *D. longispina,* and *D. cucullata* and their interspecific hybrids. Previously used genetic markers in the DLSC were not able to distinguish genome-wide signatures of introgression or the number of generations since hybridization. We employed a combination of bioinformatic approaches to investigate the history of hybridization in **Chapter 3** and **4**.

We identified the presence of *D. galeata* × *D. longispina* hybrids in population EIC and the presence of *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata* hybrids in the population SE, and ancestry paintings indicate multiple generations of hybridization and backcrossing (Figure 3.2). Hybridization between these species pairs is also commonly found in other European populations (Chin & Cristescu, 2021; Petrusek et al., 2008).

We find extensive hybridization at relatively high frequencies in two populations (EIC and SE) compared to hybrid frequencies in other taxa (see “Understanding Hybrid Zones”), and no evidence of introgression in only one population (DOB) (Figure 3.1, Figure 3.2). No direct evidence of hybridization was found for population Arendsee (AR), where only a few genomes could be sequenced, but the mitochondrial phylogeny suggests cytonuclear discordance, which has also been documented for the DLSC (Brede et al., 2009; Thielsch et al., 2017). While this can occur with incomplete lineage sorting, I propose that it is more likely the result of introgression of the mitochondrial genome following hybridization, as is frequently found in other hybrid zones (Gompert et al., 2008). Incompatibilities between the nuclear and mitochondrial genomes may also play a role in the reduced fitness of hybrids (Burton et al., 2013). Despite these patterns, the genomic divergence between the three species is clearly maintained, likely driven by ecological differentiation and intrinsic incompatibilities.

Using whole-genome data, we showed that introgression can have striking effects on the long-term evolutionary outcomes of these populations (Figure 4.3). The time-series genomic data and time since admixture estimates allowed us to untangle the decades-long history of hybridization in the two lakes. The artificially formed eutrophic lake EIC was most likely mainly colonized by *D. galeata* and *D. longispina* when the lake was constructed ~44 years ago, and hybridization is ongoing throughout this lake population. In addition, the individuals show a high degree of variation between time since admixture estimates, reflecting multiple distinct admixture events and a unique evolutionary history (Table 4.1). Interestingly, the presence of

D. galeata × *D. longispina* seems to be a recurring pattern in artificially created lakes (Griebel et al., 2015) and reservoirs (Yin et al., 2012). This could be a hint that species barriers are more disrupted in these highly anthropogenically disturbed aquatic habitats, as has been found in *Daphnia* (Millette et al., 2020) and fish (Vonlanthen et al., 2012).

Previous studies (Ma et al., 2019; Yin et al., 2014) and our findings in the resting egg bank suggest that hybrids are produced locally. However, in population EIC, the time since admixture estimates for two samples were older than the formation of the lake. This could be explained by overestimation of the junctions method or by the fact that some individuals that colonized the lake were of hybrid origin. One scenario could be that the establishment of hybrid individuals is easier in a newly formed lake with no or a very small *Daphnia* population without monopolization effects that reduce gene flow between populations (De Meester et al., 2002).

In population SE, we also found decades-old traces of hybridization in the genome of both species pairs and high variation in ancestry among individuals. Here, random dispersal and successful species establishment may play an important role (Holmes et al., 2016). For example, we found no evidence of hybridization and only *D. galeata* in lake DOB, which is located only ~4 km from SE. In lakes, the rapid population growth of *Daphnia*, local adaptation, and large resting egg banks are thought to lead to reduced success of immigrants and levels of gene flow (Brendonck & De Meester, 2003; Thielsch et al., 2009). This makes the invasion of other genotypes by the same or other species much more difficult, and only one species can establish and reproduce long-term, meaning that hybridization is unlikely in these populations.

These patterns are also supported by the recurrent findings of hybridization between these species pairs in the region over the decades (Schwenk, 1993; Spaak et al., 2004; Wolf & Mort, 1986). Interestingly, the maximum time since admixture estimates found for *D. galeata* × *D. longispina* correspond to those found for the onset of hybridization in Lake Constance and Greifensee (Brede et al., 2009). This could indicate that anthropogenic habitat disturbances in the middle of the 20th century also played a role in increasing hybridization and shifting species composition.

However, it was difficult to interpret which species and their hybrids occur in which habitats and to which ecological conditions they are linked. While we focused on the trophic level of the lakes as a well-established link to *Daphnia* species distribution, some suggest that it does not act as a strong reproductive barrier (Beninde, 2021; Chin & Cristescu, 2021) and that other ecological factors we did not consider, such as seasonal changes in phosphorus distribution (Leoni et al., 2014), food quality (Brzeziński & Von Elert, 2007), predators (Declerck & De Meester, 2003), and temperature (Spaak et al., 2012), play a more important role.

We have shown the unique features of the DLSC hybrid zone; the regular presence of up to two parental species without a collapse into a hybrid swarm, the occurrence of multiple species in mosaic aquatic hybrid zones, and the continuous hybridization and backcrossing and multiple admixture events. Time-series genomic data can also be used to distinguish between pulse hybridization and more complex ongoing hybridization (B. M. Moran et al., 2021). We find extensive admixture pulses in the genomes of individuals from both populations, but hybrid individuals were found only in the youngest sediment layers and the water column.

Hybrid fitness in general shifts between generations and habitats (Arnold & Martin, 2010), which is theorized to result in hybrids being better adapted to specific environmental conditions in an ecotone, i.e., the bounded hybrid superiority hypothesis (Moore, 1977). I propose that the temporally and spatially changing conditions in these freshwater habitats allow for the continued presence and reproduction of hybrids and parental species over time, i.e., temporal hybrid superiority (Declerck & De Meester, 2003; Spaak & Hoekstra, 1995). This appears to be the case in our populations with very recent hybrid origin and similar genomic patterns of all clones collected from the water column. For example, hybrids can potentially persist through asexual reproduction for a long time and may benefit during cold winters (Griebel et al., 2018), but the parental species have higher sexual fitness under normal conditions. This could prevent the populations from collapsing into a hybrid swarm, as gene flow across the genome is otherwise heterogeneous with few barrier regions. In addition, freshwater hybrid zones are structured by habitat and ecological gradients, such as different spawning depths in whitefish (Frei et al., 2023), which are challenging to study in our system but could also contribute to species barriers. To understand the dynamics of parental species and hybrids, further sampling of the pelagic population throughout the year and the resting egg bank is required. This could reveal whether the presence of F1 hybrids in the EIC population is a case of temporal hybrid superiority.

Overall, the maintenance of species barriers in the DLSC is likely due to post-zygotic mechanisms, such as lower sexual fitness compared to their parental species, resulting in lower production of resting eggs and lower hatching rates (Keller et al., 2007). We also found high rates of empty ephippia in our populations, which could indicate low sexual fitness. Chin et al. (2019) found that post-zygotic isolating barriers also play an important role in restricting gene flow in *D. pulex*. This could explain why hybrids are only able to dominate during certain time periods.

The chromosome-scale genome assembly allowed us to pinpoint recombination breakpoints across each chromosome with the junctions method. In populations with gene flow, genomic regions that contain genes involved in reproductive isolation or local adaptation should be more

resistant to introgression (Harrison & Larson, 2016). Introgression was ubiquitous in the genomes of individuals, with few regions indicating barriers to introgression. The presence of large homospecific ancestry regions across multiple individuals, populations, and chromosomes (Figure S4.6, Figure S4.7, Figure S4.8, Figure S4.9) confirms that hybridization is bidirectional in both *D. galeata* × *D. longispina* and *D. galeata* × *D. cucullata* hybrids, as suggested by Alric et al. (2016) and Gießler et al. (1999).

Overall, we find large variation in genome-wide ancestry among individuals with the continuous introduction of chromosomes of recent hybrid or parental origin in both species pairs and in both populations. This ranges from early hybrids living in the water column with largely heterozygous chromosomes to later-generation hybrids with mosaic genomic landscapes. Most studies combine data for whole populations and do not consider genome-wide ancestry patterns of individuals. Simulations show that individual junction patterns are strongly linked to demographic history and could tell us more about the intensity and spatial patterns of gene flow in hybrid zones (Frayer & Payseur, 2021).

We show that in the DLSC mosaic hybrid zones are still maintained despite extensive hybridization over several decades. We find genome-wide introgression with no evidence for genome stabilization. In these early generations of hybridization, the genomes are still characterized by large amounts of variation in ancestry, which can be broken up over many generations and eventually lost or fixed in all individuals (Abbott et al., 2016; Martin & Jiggins, 2017).

Implications for anthropogenic hybridization

The impact of human-mediated changes on hybridization has been documented across a range of taxa in plants, birds, insects, and mammals (Ottenburghs, 2021). These changes to ecosystems, such as eutrophication (Vonlanthen et al., 2012), climate change (Chunco, 2014; Taylor et al., 2015), and direct human-mediated translocation of species, are altering habitats and ecological gradients and create new opportunities for hybridization (Grabenstein & Taylor, 2018). By disrupting barriers between species, hybridization often persists over multiple generations and results in introgression, depending on hybrid fitness and hybridization rate. Common hybrid zone scenarios include the production of inviable or sterile F1 hybrids, bimodal hybrid zones, or a hybrid swarm (McFarlane & Pemberton, 2019).

Hybrid zones that have been affected by anthropogenic change can be studied as replicate hybrid zones because disturbances often occur in multiple locations and at different time scales (McFarlane & Pemberton, 2019). This is an important aspect of understanding the repeatability of genomic architecture (Westram et al., 2021).

While hybridization can impair local adaptation by breaking up co-adapted alleles (Runemark et al., 2019), it can also enhance genetic variation that contributes to local adaptation. Local introgression can be widespread and highly heterogeneous, with few loci exhibiting asymmetric introgression, which indicates reproductive isolation or adaptive potential, and the majority of loci are driven by genetic drift rather than selection (McFarlane et al., 2021). Introgression is known to contribute to localized variation in traits in cavefish (R. L. Moran et al., 2022) and can have long-term effects, e.g., introgression from an extinct whitefish species aided local adaptation to deep water (Frei et al., 2023). Direct evidence for adaptive introgression following anthropogenic change is rare, but could prove problematic in the future by further facilitating the success of invasive species (Valencia-Montoya et al., 2020). In mosaic aquatic hybrid zones, this has been documented for *Mytilus* (Fraïsse et al., 2014) and *Ciona* (Fraïsse et al., 2022).

Rapid local adaptation has been found in *D. pulicaria* in response to anthropogenic salinization (Wersebe & Weider, 2023), and in *D. magna* (A. Chaturvedi et al., 2021) and the DLSC (Ravindran et al., 2019; Tams et al., 2018). *Daphnia* are able to persist in habitats affected by eutrophication, such as population EIC (**Chapter 4**) and populations in peri-Alpine lakes (Alric et al., 2016; Spaak et al., 2012), or by salinization (Wersebe & Weider, 2023). Direct anthropogenic introduction is also an important process in *Daphnia* through the unintentional movement of resting eggs to new habitats (Duffy et al., 2000; Frisch et al., 2013). I suggest that these factors increase the likelihood that anthropogenic hybridization in the DLSC has the potential to lead to adaptive introgression.

In contrast to the naturally oligotrophic peri-Alpine lakes with displacement of the native *D. longispina* populations by *D. galeata* and *D. galeata* × *D. longispina* hybrids (Alric et al., 2016; Brede et al., 2009; Möst, 2013; Rellstab et al., 2011), we find two distinct patterns of species distribution in our hybrid populations. In the sampled genomes of the eutrophic population EIC, we find the predominant presence of *D. galeata* in the first decades of the lake's history, followed by continuous hybridization and an increased level of backcrossing with *D. longispina*. For population SE, the presence of *D. galeata* and two hybrid species pairs hints that random dispersal and successful species establishment play a role here. This highlights the need to study these hybrid zones across multiple spatial and temporal scales.

One of the questions anthropogenic hybridization poses is: What is a hybrid and what is a species? A parental species is usually defined as having a hybrid index of exactly 0 or 1. Small genotyping errors or only partially fixed ancestry informative markers can disrupt this, as we found for the DLSC in **Chapter 3** and **4**. This makes the definition and identification of parental species and hybrids challenging, especially in populations with many generations of

backcrossing and no historical record of species introduction. This is a concern for accurately detecting the number of generations since hybridization for conservation efforts (VonHoldt et al., 2018), e.g., in wolves (Donfrancesco et al., 2019).

Our results provide a way to differentiate this more clearly by using the junctions method, which can be applied to genomic data from different species, to understand ongoing hybridization events and illustrate the importance of identifying recurrent hybridization and heterogeneous chromosomes that could further disrupt species barriers.

Future Perspectives

The work presented in this thesis can serve as the foundation to further our understanding of the genome-wide consequences of hybridization in the DLSC and mosaic hybrid zones in general and raises new evolutionary questions.

My thesis includes the first population genomic investigation of the DLSC using the resting egg bank. Despite this advance, the sediment conditions varied from lake to lake, and poor sediment quality can inhibit whole-genome sequencing or make it challenging to date sediment cores. In addition, some lakes had lower egg densities, higher percentages of empty ephippia, or lower success rates with the WGA protocol. These limitations are difficult to predict and limited from which and how many lakes we were able to sequence samples from and proceed with time-series genomic sampling. Additional spatial and temporal sampling could further confirm how common ongoing hybridization is in the DLSC and whether the genomic patterns are similar across populations.

An easy improvement to the pre-sequencing contamination screening presented in **Chapter 2** would be to expand the PCR markers used from one to multiple mitochondrial and nuclear markers to increase the sensitivity of detecting *Daphnia* reads. Another improvement would be to use real-time PCR to quantify the amount of *Daphnia* DNA before sequencing to increase the overall success rate of amplification of resting eggs. Resurrection studies are often constrained by the fact that egg viability and hatching success decrease with increasing age (Burge et al., 2018). This may also be an issue with degraded DNA from older sediment layers, similar to that found in historical DNA (Raxworthy & Smith, 2021). Furthermore, O'Grady et al. (2022) found that WGA kits from different companies varied in their performance for amplifying sedimentary resting eggs. Combined with our findings, an ideal scenario would include testing different protocols for a specific population before proceeding with the genomic investigation.

The *D. galeata* reference genome established in **Chapter 3** and **4** showed high mapping success for the other two species and no reference bias for SNP calling. Nevertheless, reference genome assemblies for *D. longispina* and *D. cucullata* could help resolve fine-scale

differences such as structural variation (Mérot et al., 2023). The use of linked-read and long-read sequencing to analyze haplotype information is also a promising avenue to further understand the temporal dynamics of hybridization events (Leitwein et al., 2020).

The genome annotation we provide can be used to identify the biological function of introgressed genomic regions or highly divergent regions between species that underlie adaptation to different ecological conditions. Genomic cline analysis of the data set could further differentiate if alleles that introgress are neutral, advantageous, or barrier loci (Payseur, 2010). In addition, species-specific recombination rates for the DLSC could help to further improve the accuracy of time since admixture estimates from junctions and test whether introgression and recombination rates are correlated, as discussed in the general introduction.

Hybridization between species can sometimes be observed on both a contemporary and ancient scale (S. Chaturvedi et al., 2020), and more work is needed to distinguish between them in the genome. Potentially, this may also be the case in the DLSC (Beninde, 2021), which could make our ancestry informative sites from reference species less accurate and interfere with the analysis of our populations if they were affected by ancient introgression.

We were able to identify hybridization in two lakes with different ecological conditions. However, sampling of more lakes with very similar conditions in this region (e.g., a eutrophic artificial lake), as well as ecosystems that were not studied with different predators, temperature, and oligotrophic conditions, is necessary to understand these patterns in the DLSC. In addition, lake stratification and seasonal fluctuations in phosphorus could also play a role that we were not able to capture.

While the majority of studies have focused on the three most common species in the DLSC, other species are also known to hybridize (Ma et al., 2019), or I suspect that they have the potential to hybridize because we know so little about them. Adding to the limited genomic data (Beninde, 2021) for these species could answer these questions about potentially unknown hybridization, local adaptation, or even hybrid species.

Sampling efforts for all taxa should focus on replicated hybrid zones and transects across hybrid zones, especially in complex hybrid zones, (e.g., Grummer et al., 2021; Westram et al., 2021), as well as temporal sampling in potentially changing hybrid zones (Buggs, 2007). This is necessary to characterize if the genomic basis of local adaptation and reproductive isolation is universal (Harrison & Larson, 2016).

Hybridization in the DLSC is of particular importance because freshwater ecosystems are under threat from increasing anthropogenic changes, such as climate change and eutrophication, which may alter existing hybrid zones in this system and other taxa

(Grabenstein & Taylor, 2018). To resolve whether this is the case for *Daphnia* hybrid zones, we can compare hybridization patterns between more and less disturbed habitats (Grabenstein & Taylor, 2018). In addition, we can test these assumptions in natural large-scale experiments in freshwater systems, such as heated lakes (Dziuba et al., 2020), to test future outcomes. *Daphnia* is also one of the few hybrid systems where experimental hybridization over multiple generations in the laboratory or mesocosms would be feasible to understand the long-term fitness of hybrids and genomic consequences (e.g., Franchini et al., 2018; Pritchard et al., 2013). Taken together, this could help us to understand why hybridization is common in *Daphnia*, how it is linked to anthropogenic change, and how species barriers are maintained now and in the future.

One aspect we were not able to examine is the possibility of introgression happening via a conduit species, which allows introgression between two species that do not directly overlap (Grant & Grant, 2020). Another possibility is hybridization between more than two species (Natola et al., 2022; Toews et al., 2018). In our system, there could be undetected hybridization between *D. longispina* and *D. cucullata*, which are generally separated by strong ecological barriers but can occur in the same population. These scenarios are still rarely studied and few genomic tools are available to analyze hybridization with ancestry from three species (Banerjee et al., 2023).

Understanding the magnitude and timing of gene flow in populations is an important step in understanding the genomic outcomes of hybridization (Payseur & Rieseberg, 2016). Finally, our findings raise the possibility that recurrent hybridization pulses and large variation in genome-wide ancestry among individuals are not limited to *Daphnia*, and that complex hybridization dynamics need to be considered when studying recent hybrid zones. Moving forward, these directions could elucidate how genetic and evolutionary processes shape the genomes in hybrid zones in the past and present.

Conclusion

This thesis reveals the evolutionary complexity of hybridization and introgression in the DLSC and mosaic hybrid zones. By providing methods for genomic investigation of the *Daphnia* resting egg bank and a chromosome-scale genome assembly, we were able to identify extensive introgression and hybridization across the genome in the DLSC. A more detailed analysis of the hybridization dynamics in these populations uncovered successive hybridization events over several decades that led to diverse genomic landscapes of hybrid individuals and hybrid zones with unique characteristics. The extended theory of junctions enables us to accurately date the time since admixture and illustrates its use for other species with a history of hybridization.

Combined, the work presented in this thesis establishes the DLSC as an emerging hybrid system with unique characteristics and the ability to study past hybridization events. I expect hybridization to continue in these populations, and it is not yet clear if these hybrid zones are structured by stable habitat and ecological gradients, or if this will lead to negative evolutionary consequences such as species collapse that is potentially exacerbated by increasing anthropogenic changes in freshwater ecosystems. Open questions remain about which genome properties are responsible for these patterns, how species boundaries are maintained in the face of gene flow, and the genomic basis of adaptation.

In a broader context, this work gives us insight into other complex hybrid zones and the different aspects that shape the genomic landscape, such as hybridization between multiple species, movement of hybrid zones, and recurrent hybridization pulses over long periods of time.

Discussion References

Abbott, R. J., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R. J., Barton, N. H., & Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, *25*(11), 2325–2332. https://doi.org/10.1111/mec.13685

Aguillon, S. M., & Rohwer, V. G. (2022). Revisiting a classic hybrid zone: Movement of the northern flicker hybrid zone in contemporary times. *Evolution*, *76*(5), 1082–1090. https://doi.org/10.1111/evo.14474

Alric, B., Möst, M., Domaizon, I., Pignol, C., Spaak, P., & Perga, M.-E. (2016). Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology*, *29*(4), 720–735. https://doi.org/10.1111/jeb.12820

Altshuler, I., Demiri, B., Xu, S., Constantin, A., Yan, N. D., & Cristescu, M. E. (2011). An Integrated Multi-Disciplinary Approach for Studying Multiple Stressors in Freshwater Ecosystems: *Daphnia* as a Model Organism. *Integrative and Comparative Biology*, *51*(4), 623–633. https://doi.org/10.1093/icb/icr103

Arnold, M. L., & Martin, N. H. (2010). Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*, *25*(9), 530–536. https://doi.org/10.1016/j.tree.2010.06.005

Banerjee, S. M., Powell, D. L., Moran, B. M., Ramírez-Duarte, W. F., Langdon, Q. K., Gunn, T. R., Vazquez, G., Rochman, C., & Schumer, M. (2023). Complex hybridization between deeply diverged fish species in a disturbed ecosystem. *Evolution*, *77*(4), 995–1005. https://doi.org/10.1093/evolut/qpad019

Barnard-Kubow, K. B., Becker, D., Murray, C. S., Porter, R., Gutierrez, G., Erickson, P., Nunez, J. C. B., Voss, E., Suryamohan, K., Ratan, A., Beckerman, A., & Bergland, A. O. (2022). Genetic Variation in Reproductive Investment Across an Ephemerality Gradient in *Daphnia pulex*. *Molecular Biology and Evolution*, *39*(6), msac121. https://doi.org/10.1093/molbev/msac121

Barth, J. M. I., Gubili, C., Matschiner, M., Tørresen, O. K., Watanabe, S., Egger, B., Han, Y.-S., Feunteun, E., Sommaruga, R., Jehle, R., & Schabetsberger, R. (2020). Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-15099-x

Beninde, J. (2021). *Evolution of the Daphnia longispina Group in Response to Anthropogenic Environmental Changes* [Ph.D. thesis, Universität Konstanz]. https://kops.uni-konstanz.de/handle/123456789/54238

Beninde, J., Möst, M., & Meyer, A. (2020). Optimized and affordable high-throughput sequencing workflow for preserved and nonpreserved small zooplankton specimens. *Molecular Ecology Resources*, *20*(6), 1632–1646. https://doi.org/10.1111/1755-0998.13228

Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., & Schwenk, K. (2009). The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences*, *106*(12), 4758–4763. https://doi.org/10.1073/pnas.0807187106

Brendonck, L., & De Meester, L. (2003). Egg banks in freshwater zooplankton: Evolutionary and ecological archives in the sediment. *Hydrobiologia*, *491*(1), 65–84. https://doi.org/10.1023/A:1024454905119

Brzeziński, T., & Von Elert, E. (2007). Biochemical food quality effects on a *Daphnia* hybrid complex. *Limnology and Oceanography*, *52*(6), 2350–2357. https://doi.org/10.4319/lo.2007.52.6.2350

Buggs, R. J. A. (2007). Empirical study of hybrid zone movement. *Heredity*, *99*(3), 301–312. https://doi.org/10.1038/sj.hdy.6800997

Burge, D. R. L., Edlund, M. B., & Frisch, D. (2018). Paleolimnology and resurrection ecology: The future of reconstructing the past. *Evolutionary Applications*, *11*(1), 42–59. https://doi.org/10.1111/eva.12556

Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 281–302. https://doi.org/10.1146/annurev-ecolsys-110512-135758

Calfee, E., Agra, M. N., Palacio, M. A., Ramírez, S. R., & Coop, G. (2020). Selection and hybridization shaped the rapid spread of African honey bee ancestry in the Americas. *PLoS Genetics*, *16*(10), e1009038. https://doi.org/10.1371/journal.pgen.1009038

Chaturvedi, A., Zhou, J., Raeymaekers, J. A. M., Czypionka, T., Orsini, L., Jackson, C. E., Spanier, K. I., Shaw, J. R., Colbourne, J. K., & De Meester, L. (2021). Extensive standing genetic variation from a small number of founders enables rapid adaptation in *Daphnia*. *Nature Communications*, *12*(1), 4306. https://doi.org/10.1038/s41467-021-24581-z

Chaturvedi, S., Lucas, L. K., Buerkle, C. A., Fordyce, J. A., Forister, M. L., Nice, C. C., & Gompert, Z. (2020). Recent hybrids recapitulate ancient hybrid outcomes. *Nature Communications*, *11*(1), 2179. https://doi.org/10.1038/s41467-020-15641-x

Chhatre, V. E., Evans, L. M., DiFazio, S. P., & Keller, S. R. (2018). Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Molecular Ecology*, *27*(23), 4820–4838. https://doi.org/10.1111/mec.14820

Chin, T. A., Cáceres, C. E., & Cristescu, M. E. (2019). The evolution of reproductive isolation in *Daphnia*. *BMC Evolutionary Biology*, *19*(1), 216. https://doi.org/10.1186/s12862-019-1542-9

Chin, T. A., & Cristescu, M. E. (2021). Speciation in *Daphnia*. *Molecular Ecology*, *30*(6), 1398–1418. https://doi.org/10.1111/mec.15824

Chunco, A. J. (2014). Hybridization in a warmer world. *Ecology and Evolution*, *4*(10), 2019–2031. https://doi.org/10.1002/ece3.1052

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., Sawkins, A., Snead, A. A., Whalen, J. C., & Pinsky, M. L. (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13789

Cordeiro, E. M. G., Pantoja-Gomez, L. M., De Paiva, J. B., Nascimento, A. R. B., Omoto, C., Michel, A. P., & Correa, A. S. (2020). Hybridization and introgression between *Helicoverpa armigera* and *H. zea*: An adaptational bridge. *BMC Evolutionary Biology*, *20*(1), 61. https://doi.org/10.1186/s12862-020-01621-8

De Meester, L., Gómez, A., Okamura, B., & Schwenk, K. (2002). The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, *23*(3), 121–135. https://doi.org/10.1016/S1146-609X(02)01145-1

Declerck, S., & De Meester, L. (2003). Impact of fish predation on coexisting *Daphnia* taxa: A partial test of the temporal hybrid superiority hypothesis. *Hydrobiologia*, *500*(1), 83–94. https://doi.org/10.1023/A:1024656602248

De-Kayne, R., Selz, O. M., Marques, D. A., Frei, D., Seehausen, O., & Feulner, P. G. D. (2022). Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. *Nature Communications*, *13*(1), 4479. https://doi.org/10.1038/s41467-022-32181-8

Donfrancesco, V., Ciucci, P., Salvatori, V., Benson, D., Andersen, L. W., Bassi, E., Blanco, J. C., Boitani, L., Caniglia, R., Canu, A., Capitani, C., Chapron, G., Czarnomska, S. D., Fabbri, E., Galaverni, M., Galov, A., Gimenez, O., Godinho, R., Greco, C., … Mukherjee, N. (2019). Unravelling the Scientific Debate on How to Address Wolf-Dog Hybridization in Europe. *Frontiers in Ecology and Evolution*, *7*, 175. https://doi.org/10.3389/fevo.2019.00175

Duffy, M. A., Perry, L. J., Kearns, C. M., Weider, L. J., & Hairston, N. G. (2000). Paleogenetic evidence for a past invasion of Onondaga Lake, New York, by exotic *Daphnia curvirostris* using mtDNA from dormant eggs. *Limnology and Oceanography*, *45*(6), 1409–1414. https://doi.org/10.4319/lo.2000.45.6.1409

Duranton, M., Bonhomme, F., & Gagnaire, P.-A. (2019). The spatial scale of dispersal revealed by admixture tracts. *Evolutionary Applications*, *12*(9), 1743–1756. https://doi.org/10.1111/eva.12829

Dziuba, M. K., Herdegen-Radwan, M., Pluta, E., Wejnerowski, Ł., Szczuciński, W., & Cerbin, S. (2020). Temperature increase altered *Daphnia* community structure in artificially heated lakes: A potential scenario for a warmer future. *Scientific Reports*, *10*(1), 13956. https://doi.org/10.1038/s41598-020-70294-6

Fraïsse, C., Le Moan, A., Roux, C., Dubois, G., Daguin-Thiebaut, C., Gagnaire, P.-A., Viard, F., & Bierne, N. (2022). Introgression between highly divergent sea squirt genomes: An adaptive breakthrough? *Peer Community Journal*, *2*, e54. https://doi.org/10.24072/pcjournal.172

Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene-flow in a mosaic hybrid zone: Is local introgression adaptive? *Genetics*, *197*(3), 939–951. https://doi.org/10.1534/genetics.114.161380

Franchini, P., Jones, J. C., Xiong, P., Kneitz, S., Gompert, Z., Warren, W. C., Walter, R. B., Meyer, A., & Schartl, M. (2018). Long-term experimental hybridisation results in the evolution of a new sex chromosome in swordtail fish. *Nature Communications*, *9*(1), 5136. https://doi.org/10.1038/s41467-018-07648-2

Frayer, M. E., & Payseur, B. A. (2021). Demographic history shapes genomic ancestry in hybrid zones. *Ecology and Evolution*, *11*(15), 10290–10302. https://doi.org/10.1002/ece3.7833

Frei, D., Reichlin, P., Seehausen, O., & Feulner, P. G. D. (2023). Introgression from extinct species facilitates adaptation to its vacated niche. *Molecular Ecology*, *32*(4), 841–853. https://doi.org/10.1111/mec.16791

Frisch, D., Havel, J. E., & Weider, L. J. (2013). The invasion history of the exotic freshwater zooplankter *Daphnia lumholtzi* (Cladocera, Crustacea) in North America: A genetic analysis. *Biological Invasions*, *15*(4), 817–828. https://doi.org/10.1007/s10530-012-0329-3

Frisch, D., Morton, P. K., Chowdhury, P. R., Culver, B. W., Colbourne, J. K., Weider, L. J., & Jeyasingh, P. D. (2014). A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters*, *17*(3), 360–368. https://doi.org/10.1111/ele.12237

Galaverni, M., Caniglia, R., Pagani, L., Fabbri, E., Boattini, A., & Randi, E. (2017). Disentangling Timing of Admixture, Patterns of Introgression, and Phenotypic Indicators in a Hybridizing Wolf Population. *Molecular Biology and Evolution*, *34*(9), 2324–2339. https://doi.org/10.1093/molbev/msx169

Gießler, S., Mader, E., & Schwenk, K. (1999). Morphological evolution and genetic differentiation in *Daphnia* species complexes. *Journal of Evolutionary Biology*, *12*(4), 710–723. https://doi.org/10.1046/j.1420-9101.1999.00065.x

Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, *17*(24), 5231–5244. https://doi.org/10.1111/j.1365-294X.2008.03988.x

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking Barriers: Causes, Consequences, and Experimental Utility of Human-Mediated Hybridization. *Trends in Ecology & Evolution*, *33*(3), 198–212. https://doi.org/10.1016/j.tree.2017.12.008

Grant, P. R., & Grant, B. R. (2020). Triad hybridization via a conduit species. *Proceedings of the National Academy of Sciences*, *117*(14), 7888–7896. https://doi.org/10.1073/pnas.2000388117

Griebel, J., Gießler, S., Poxleitner, M., Faria, A. N., Yin, M., & Wolinska, J. (2015). Extreme Environments Facilitate Hybrid Superiority – The Story of a Successful *Daphnia galeata* × *longispina* Hybrid Clone. *PLoS One*, *10*(10), e0140275. https://doi.org/10.1371/journal.pone.0140275

Griebel, J., Utz, M., Hermisson, J., & Wolinska, J. (2018). The establishment of hybrids of the *Daphnia longispina* complex explained by a mathematical model incorporating different overwintering life history strategies. *PLoS One*, *13*(7), e0200802. https://doi.org/10.1371/journal.pone.0200802

Grummer, J. A., Avila, L. J., Morando, M. M., & Leaché, A. D. (2021). Four Species Linked by Three Hybrid Zones: Two Instances of Repeated Hybridization in One Species Group (Genus *Liolaemus*). *Frontiers in Ecology and Evolution*, *9*, 624109. https://doi.org/0.3389/fevo.2021.624109

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454–2466. https://doi.org/10.1111/mec.13582

Holmes, C. J., Pantel, J. H., Schulz, K. L., & Cáceres, C. E. (2016). Initial genetic diversity enhances population establishment and alters genetic structuring of a newly established *Daphnia* metapopulation. *Molecular Ecology*, *25*(14), 3299–3308. https://doi.org/10.1111/mec.13672

Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, *30*(17), 4245–4258. https://doi.org/10.1111/mec.16056

Jankowski, T., & Straile, D. (2003). A comparison of egg-bank and long-term plankton dynamics of two *Daphnia* species, *D. hyalina* and *D. galeata*: Potentials and limits of reconstruction. *Limnology and Oceanography*, *48*(5), 1948–1955. https://doi.org/10.4319/lo.2003.48.5.1948

Janzen, T., & Miró Pina, V. (2022). Estimating the time since admixture from phased and unphased molecular data. *Molecular Ecology Resources*, *22*(3), 908–926. https://doi.org/10.1111/1755-0998.13519

Janzen, T., Nolte, A. W., & Traulsen, A. (2018). The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. *Evolution*, *72*(4), 735–750. https://doi.org/10.1111/evo.13436

Johannesson, K., Le Moan, A., Perini, S., & André, C. (2020). A Darwinian Laboratory of Multiple Contact Zones. *Trends in Ecology & Evolution*, *35*(11), 1021–1036. https://doi.org/10.1016/j.tree.2020.07.015

Keller, B., Wolinska, J., Manca, M., & Spaak, P. (2008). Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2943–2952. https://doi.org/10.1098/rstb.2008.0044

Keller, B., Wolinska, J., Tellenbach, C., & Spaak, P. (2007). Reproductive isolation keeps hybridizing *Daphnia* species distinct. *Limnology and Oceanography*, *52*(3), 984–991. https://doi.org/10.4319/lo.2007.52.3.0984

Lack, J. B., Weider, L. J., & Jeyasingh, P. D. (2018). Whole genome amplification and sequencing of a *Daphnia* resting egg. *Molecular Ecology Resources*, *18*(1), 118–127. https://doi.org/10.1111/1755-0998.12720

Langdon, Q. K., Powell, D. L., Kim, B., Banerjee, S. M., Payne, C., Dodge, T. O., Moran, B., Fascinetto-Zago, P., & Schumer, M. (2022). Predictability and parallelism in the contemporary evolution of hybrid genomes. *PLoS Genetics*, *18*(1), e1009914. https://doi.org/10.1371/journal.pgen.1009914

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, *9*(6), 3470–3490. https://doi.org/10.1002/ece3.4981

Le Moan, A., Roby, C., Fraïsse, C., Daguin-Thiébaut, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, *30*(24), 6718–6732. https://doi.org/10.1111/mec.16189

Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, *210*, 69–84. https://doi.org/10.1016/j.aquatox.2019.02.009

Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., & Bernatchez, L. (2020). Using Haplotype Information for Conservation Genomics. *Trends in Ecology & Evolution*, *35*(3), 245–258. https://doi.org/10.1016/j.tree.2019.10.012

Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, *27*(17), 3466–3483. https://doi.org/10.1111/mec.14816

Leoni, B., Garibaldi, L., & Gulati, R. D. (2014). How does interannual trophic variability caused by vertical water mixing affect reproduction and population density of the *Daphnia longispina* group in Lake Iseo, a deep stratified lake in Italy? *Inland Waters*, *4*(2), 193–203. https://doi.org/10.5268/IW-4.2.663

Lewanski, A. L., Golcher-Benavides, J., Rick, J. A., & Wagner, C. E. (2022). Variable hybridization between two Lake Tanganyikan cichlid species in recent secondary contact. *Molecular Ecology*, *31*(19), 5041–5059. https://doi.org/10.1111/mec.16636

Ma, X., Hu, W., Smilauer, P., Yin, M., & Wolinska, J. (2019). *Daphnia galeata* and *D. dentifera* are geographically and ecologically separated whereas their hybrids occur in intermediate habitats: A survey of 44 Chinese lakes. *Molecular Ecology*, *28*(4), 785–802. https://doi.org/10.1111/mec.14991

Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, *47*, 69–74. https://doi.org/10.1016/j.gde.2017.08.007

Mattucci, F., Galaverni, M., Lyons, L. A., Alves, P. C., Randi, E., Velli, E., Pagani, L., & Caniglia, R. (2019). Genomic approaches to identify hybrids and estimate admixture times in European wildcat populations. *Scientific Reports*, *9*(1), 11612. https://doi.org/10.1038/s41598-019-48002-w

McFarlane, S. E., Hunter, D. C., Senn, H. V., Smith, S. L., Holland, R., Huisman, J., & Pemberton, J. M. (2020). Increased genetic marker density reveals high levels of admixture between red deer and introduced Japanese sika in Kintyre, Scotland. *Evolutionary Applications*, *13*(2), 432–441. https://doi.org/10.1111/eva.12880

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

McFarlane, S. E., Senn, H. V., Smith, S. L., & Pemberton, J. M. (2021). Locus-specific introgression in young hybrid swarms: Drift may dominate selection. *Molecular Ecology*, *30*(9), 2104–2115. https://doi.org/10.1111/mec.15862

Mérot, C., Stenløkk, K. S. R., Venney, C., Laporte, M., Moser, M., Normandeau, E., Árnyasi, M., Kent, M., Rougeux, C., Flynn, J. M., Lien, S., & Bernatchez, L. (2023). Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (*Coregonus* sp.) with long and short reads. *Molecular Ecology*, *32*(6), 1458–1477. https://doi.org/10.1111/mec.16468

Millette, K. L., Gonzalez, A., & Cristescu, M. E. (2020). Breaking ecological barriers: Anthropogenic disturbance leads to habitat transitions, hybridization, and high genetic diversity. *Science of The Total Environment*, *740*, 140046. https://doi.org/10.1016/j.scitotenv.2020.140046

Moore, W. S. (1977). An Evaluation of Narrow Hybrid Zones in Vertebrates. *The Quarterly Review of Biology*, *52*(3), 263–277. https://doi.org/10.1086/409995

Moran, B. M., Payne, C., Langdon, Q., Powell, D. L., Brandvain, Y., & Schumer, M. (2021). The genomic consequences of hybridization. *eLife*, *10*, e69016. https://doi.org/10.7554/eLife.69016

Moran, R. L., Jaggard, J. B., Roback, E. Y., Kenzior, A., Rohner, N., Kowalko, J. E., Ornelas-García, C. P., McGaugh, S. E., & Keene, A. C. (2022). Hybridization underlies localized trait evolution in cavefish. *iScience*, *25*(2), 103778. https://doi.org/10.1016/j.isci.2022.103778

Möst, M. (2013). *Environmental change and its impact on hybridising Daphnia species complexes* [Ph.D. thesis, ETH Zürich]. https://doi.org/10.3929/ethz-a-010076219

Natola, L., Seneviratne, S. S., & Irwin, D. (2022). Population genomics of an emergent tri-species hybrid zone. *Molecular Ecology*, *31*(20), 5356–5367. https://doi.org/10.1111/mec.16650

Nelson, J. R. (2014). Random-Primed, Phi29 DNA Polymerase-Based Whole Genome Amplification. *Current Protocols in Molecular Biology*, *105*(1), 15.13.1–15.13.16. https://doi.org/10.1002/0471142727.mb1513s105

Nouhaud, P., Martin, S. H., Portinha, B., Sousa, V. C., & Kulmuni, J. (2022). Rapid and predictable genome evolution across three hybrid ant populations. *PLoS Biology*, *20*(12), e3001914. https://doi.org/10.1371/journal.pbio.3001914

O'Grady, C. J., Dhandapani, V., Colbourne, J. K., & Frisch, D. (2022). Refining the evolutionary time machine: An assessment of whole genome amplification using single historical *Daphnia* eggs. *Molecular Ecology Resources*, *22*(3), 946–961. https://doi.org/10.1111/1755-0998.13524

Orsini, L., Spanier, K. I., & De Meester, L. (2012). Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: Validation in space, time and experimental evolution. *Molecular Ecology*, *21*(9), 2160–2175. https://doi.org/10.1111/j.1365-294X.2011.05429.x

Ottenburghs, J. (2019). Multispecies hybridization in birds. *Avian Research*, *10*(1), 20. https://doi.org/10.1186/s40657-019-0159-4

Ottenburghs, J. (2021). The genic view of hybridization in the Anthropocene. *Evolutionary Applications*, *14*(10), 2342–2360. https://doi.org/10.1111/eva.13223

Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, *10*(5), 806–820. https://doi.org/10.1111/j.1755-0998.2010.02883.x

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Petrusek, A., Seda, J., Macháček, J., Ruthová, Š., & Šmilauer, P. (2008). *Daphnia* hybridization along ecological gradients in pelagic environments: The potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1505), 2931–2941. https://doi.org/10.1098/rstb.2008.0026

Pfenninger, M., Doria, H. B., Nickel, J., Thielsch, A., Schwenk, K., & Cordellier, M. (2022). Spontaneous rate of clonal single nucleotide mutations in *Daphnia galeata*. *PLoS One*, *17*(4), e0265632. https://doi.org/10.1371/journal.pone.0265632

Pritchard, V. L., Knutson, V. L., Lee, M., Zieba, J., & Edmands, S. (2013). Fitness and morphological outcomes of many generations of hybridization in the copepod *Tigriopus californicus*. *Journal of Evolutionary Biology*, *26*(2), 416–433. https://doi.org/10.1111/jeb.12060

Ravindran, S. P., Herrmann, M., & Cordellier, M. (2019). Contrasting patterns of divergence at the regulatory and sequence level in European *Daphnia galeata* natural populations. *Ecology and Evolution*, *9*(5), 2487–2504. https://doi.org/10.1002/ece3.4894

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: Advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Rellstab, C., Keller, B., Girardclos, S., Anselmetti, F. S., & Spaak, P. (2011). Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. *Limnology and Oceanography*, *56*(1), 292–302. https://doi.org/10.4319/lo.2011.56.1.0292

Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Béduneau, V., Briard, O., El Ayari, T., Hochscheid, S., Belkhir, K., Arnaud-Haond, S., Gagnaire, P.-A., & Bierne, N. (2019). Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. *Evolution*, *73*(4), 817–835. https://doi.org/10.1111/evo.13696

Rougemont, Q., Perrier, C., Besnard, A.-L., Lebel, I., Abdallah, Y., Feunteun, E., Réveillac, E., Lasne, E., Acou, A., Nachón, D. J., Cobo, F., Evanno, G., Baglinière, J.-L., & Launey, S. (2022). Population genetics reveals divergent lineages and ongoing hybridization in a declining migratory fish species complex. *Heredity*, *129*(2), 137–151. https://doi.org/10.1038/s41437-022-00547-9

Runemark, A., Vallejo-Marin, M., & Meier, J. I. (2019). Eukaryote hybrid genomes. *PLoS Genetics*, *15*(11), e1008404. https://doi.org/10.1371/journal.pgen.1008404

Ryan, S. F., Deines, J. M., Scriber, J. M., Pfrender, M. E., Jones, S. E., Emrich, S. J., & Hellmann, J. J. (2018). Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proceedings of the National Academy of Sciences*, *115*(10), E2284–E2291. https://doi.org/10.1073/pnas.1714950115

Satokangas, I., Nouhaud, P., Seifert, B., Punttila, P., Schultz, R., Jones, M. M., Sirén, J., Helanterä, H., & Kulmuni, J. (2023). Semipermeable species boundaries create opportunities for gene flow and adaptive potential. *Molecular Ecology*, *32*(15), 4329–4347. https://doi.org/10.1111/mec.16992

Schwenk, K. (1993). Interspecific hybridization in *Daphnia*: Distinction and origin of hybrid matrilines. *Molecular Biology and Evolution*, *10*(6), 1289–1302. https://doi.org/10.1093/oxfordjournals.molbev.a040076

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Simon, A., Arbiol, C., Nielsen, E. E., Couteau, J., Sussarellu, R., Burgeot, T., Bernard, I., Coolen, J. W. P., Lamy, J.-B., Robert, S., Skazina, M., Strelkov, P., Queiroga, H., Cancio, I., Welch, J. J., Viard, F., & Bierne, N. (2020). Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. *Evolutionary Applications*, *13*(3), 575–599. https://doi.org/10.1111/eva.12879

Smith, K. L., Hale, J. M., Gay, L., Kearney, M., Austin, J. J., Parris, K. M., & Melville, J. (2013). Spatio-temporal changes in the structure of an Australian frog hybrid zone: A 40-year perspective. *Evolution*, *67*(12), 3442–3454. https://doi.org/10.1111/evo.12140

Spaak, P., Denk, A., Boersma, M., & Weider, L. J. (2004). Spatial and temporal patterns of sexual reproduction in a hybrid *Daphnia* species complex. *Journal of Plankton Research*, *26*(6), 625–635. https://doi.org/10.1093/plankt/fbh064

Spaak, P., Fox, J., & Hairston, N. G. (2012). Modes and mechanisms of a *Daphnia* invasion. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2936–2944. https://doi.org/10.1098/rspb.2012.0280

Spaak, P., & Hoekstra, J. R. (1995). Life History Variation and the Coexistence of a *Daphnia* Hybrid With Its Parental Species. *Ecology*, *76*(2), 553–564. https://doi.org/10.2307/1941213

Tams, V., Lüneburg, J., Seddar, L., Detampel, J.-P., & Cordellier, M. (2018). Intraspecific phenotypic variation in life history traits of *Daphnia galeata* populations in response to fish kairomones. *PeerJ*, *6*, e5746. https://doi.org/10.7717/peerj.5746

Taylor, S. A., Larson, E. L., & Harrison, R. G. (2015). Hybrid zones: Windows on climate change. *Trends in Ecology & Evolution*, *30*(7), 398–406. https://doi.org/10.1016/j.tree.2015.04.010

Thielsch, A., Brede, N., Petrusek, A., De Meester, L., & Schwenk, K. (2009). Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Molecular Ecology*, *18*(8), 1616–1628. https://doi.org/10.1111/j.1365-294X.2009.04130.x

Thielsch, A., Knell, A., Mohammadyari, A., Petrusek, A., & Schwenk, K. (2017). Divergent clades or cryptic species? Mito-nuclear discordance in a *Daphnia* species complex. *BMC Evolutionary Biology*, *17*, 227. https://doi.org/10.1186/s12862-017-1070-4

Toews, D. P. L., Streby, H. M., Burket, L., & Taylor, S. A. (2018). A wood-warbler produced through both interspecific and intergeneric hybridization. *Biology Letters*, *14*(11), 20180557. https://doi.org/10.1098/rsbl.2018.0557

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., Gordon, K. H. J., Specht, A., Paula-Moraes, S. V., Rane, R., Walsh, T. K., & Jiggins, C. D. (2020). Adaptive Introgression across Semipermeable Species Boundaries between Local *Helicoverpa zea* and Invasive *Helicoverpa armigera* Moths. *Molecular Biology and Evolution*, *37*(9), 2568–2583. https://doi.org/10.1093/molbev/msaa108

Vaníčková, I., Seda, J., & Petrusek, A. (2010). The stabilizing effect of resting egg banks of the *Daphnia longispina* species complex for longitudinal taxon heterogeneity in long and narrow reservoirs. *Hydrobiologia*, *643*(1), 85–95. https://doi.org/10.1007/s10750-010-0126-x

Veller, C., Edelman, N. B., Muralidhar, P., & Nowak, M. A. (2023). Recombination and selection against introgressed DNA. *Evolution*, *77*(4), 1131–1144. https://doi.org/10.1093/evolut/qpad021

Viard, F., Riginos, C., & Bierne, N. (2020). Anthropogenic hybridization at sea: Three evolutionary questions relevant to invasive species management. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *375*(1806), 20190547. https://doi.org/10.1098/rstb.2019.0547

VonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., & Rutledge, L. Y. (2018). Redefining the Role of Admixture and Genomics in Species Conservation. *Conservation Letters*, *11*(2), e12371. https://doi.org/10.1111/conl.12371

Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., Roy, D., Di Piazza, S., Largiader, C. R., & Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, *482*(7385), 357–362. https://doi.org/10.1038/nature10824

Wersebe, M. J., Sherman, R. E., Jeyasingh, P. D., & Weider, L. J. (2023). The roles of recombination and selection in shaping genomic divergence in an incipient ecological species complex. *Molecular Ecology*, *32*(6), 1478–1496. https://doi.org/10.1111/mec.16383

Wersebe, M. J., & Weider, L. J. (2023). Resurrection genomics provides molecular and phenotypic evidence of rapid adaptation to salinization in a keystone aquatic species. *Proceedings of the National Academy of Sciences*, *120*(6), e2217276120. https://doi.org/10.1073/pnas.2217276120

Westram, A. M., Faria, R., Johannesson, K., & Butlin, R. (2021). Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, *30*(15), 3797–3814. https://doi.org/10.1111/mec.15861

Wielstra, B. (2019). Historical hybrid zone movement: More pervasive than appreciated. *Journal of Biogeography*, *46*(7), 1300–1305. https://doi.org/10.1111/jbi.13600

Wolf, H. G., & Mort, M. A. (1986). Inter-specific hybridization underlies phenotypic variability in *Daphnia* populations. *Oecologia*, *68*(4), 507–511. https://doi.org/10.1007/BF00378763

Yin, M., Gießler, S., Griebel, J., & Wolinska, J. (2014). Hybridizing *Daphnia* communities from ten neighbouring lakes: Spatio-temporal dynamics, local processes, gene flow and invasiveness. *BMC Evolutionary Biology*, *14*(1), 80. https://doi.org/10.1186/1471-2148-14-80

Yin, M., Petrusek, A., Seda, J., & Wolinska, J. (2012). Fine-scale temporal and spatial variation of taxon and clonal structure in the *Daphnia longispina* hybrid complex in heterogeneous environments. *BMC Evolutionary Biology*, *12*(1), 12. https://doi.org/10.1186/1471-2148-12-12

Zbinden, Z. D., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2023). A community genomics approach to natural hybridization. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1999), 20230768. https://doi.org/10.1098/rspb.2023.0768

Appendix

List of Figures

List of Tables

Author Contributions

Chapter 2 – Cost-saving population genomic investigation of *Daphnia longispina* complex resting eggs using whole genome amplification and pre-sequencing screening:

J.N. and M.C. designed the study. **J.N.** performed whole genome amplification and library preparation of resting eggs and analyzed the genomic data. **J.N.** wrote the first draft of the manuscript with input from M.C.

Chapter 3 – Hybridization dynamics and extensive introgression in the *Daphnia longispina* species complex: new insights from a high-quality *Daphnia galeata* reference genome:

J.N. and M.C. conceived of the study. M.C., **J.N.**, A.T., S.R.D., and M.M. performed the sampling and wet lab work. T.S., **J.N.**, and A.T. conducted the genome assembly and annotation. **J.N.** conducted the population genomic analysis and mitochondrial genome reconstruction. T.H. and M.M. conducted the phylogenetic analyses. **J.N.** wrote the majority of the first draft with input from M.C. M.M. and T.H. contributed sections on the phylogenetic analysis, T.S. on the genome assembly and annotation, and M.P. on the F_{ST} outlier windows analysis. **All authors** edited and contributed to the final version and gave final approval for preprint deposition and publication.

Chapter 4 – Patterns of recurrent hybridization and genomic architecture through time in the *Daphnia longispina* species complex:

J.N. and M.C. designed the study and performed the sampling. **J.N.** performed whole genome amplification and library preparation of resting eggs and DNA extractions, and M.C. performed the Hi-C library preparation. T.S. conducted the Hi-C genome assembly and annotation. **J.N.** conducted the SNP calling and population genomic analysis. T.J. conducted the time since admixture estimates using junctions. **J.N.** wrote the first draft of the manuscript with input from M.M. T.S. contributed sections on Hi-C genome assembly and annotation, and T.J. contributed methods and results for junctions. **All authors** provided feedback on the final version.

Dr. Mathilde Cordellier

Supplementary Information on Chapter 2

Table S2.1. Accession number for genomes included in the custom database to assess contamination with FastQ Screen.

Genome	GenBank accession
Homo sapiens	GCA_000001405.22
UniVec	https://ftp.ncbi.nlm.nih.gov/pub/UniVec/
Bacteria	https://ftp.ncbi.nlm.nih.gov/refseq/release/bacteria/
Virus	https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/
Daphnia galeata	GCA_918697745.1
Daphnia magna	GCA_001632505.1
Acutodesmus obliquus	GCA_002149895.1

Table S2.2. Summary of sequencing statistics for all successful amplification samples that were sequenced to a higher coverage using NovaSeq. Mean genome coverage is based on the estimated genome size of 156 Mb for *D. galeata*.

Sample name	Total raw reads	Total trimmed reads	Mapped reads %	Mean genome coverage
EIC_18	14,952,714	7,897,119	94.61	7.32
EIC_16	15,450,664	6,755,689	96.55	6.52
EIC_17	15,392,966	11,433,190	96.67	11.43
EIC_11	16,210,646	5,563,028	98.06	5.55
EIC_12	19,121,834	929,386	90.95	0.60
EIC_13	14,463,272	11,446,924	97.76	11.55
EIC_14	16,917,644	10,256,933	98.26	10.39
EIC_15	15,531,160	12,295,702	97.76	12.41
M5	77,594,732	65,570,833	84.75	57.05

Table S2.3. Average genome-wide heterozygosity estimated from the proportion of heterozygous genotype calls in sliding windows across the genome.

Sample name	Average genome-wide heterozygosity
EIC_18	0.01100
EIC_16	0.01225
EIC_17	0.00613
EIC_11	0.00697
EIC_12	0.00633
EIC_13	0.00980
EIC_14	0.00598
EIC_15	0.00756
M5	0.00934

Supplementary Information on Chapter 3

Supplementary materials and methods

In the following section default options are applied, if not stated otherwise.

Genome assembly

Adapter and quality trimming

Adapter and quality trimming of Illumina reads was conducted with autotrim.pl 0.6.1 (Waldvogel et al., 2018), in combination with Trimmomatic 0.38 (Bolger et al., 2014), FastQC 0.11.7 (Andrews, 2010), and MultiQC 1.6 (Ewels et al., 2016) using a custom adapter file and the following Trimmomatic parameters: ILLUMINACLIP: adapter_combined.fa: 2:30:10 TRAILING: 15 SLIDINGWINDOW: 4:20 MINLEN: 50. The tool autotrim.pl allows to trim additionally to adapter and quality for overrepresented k-mers which probably arise from smaller parts of adapter sequences. PacBio subreads were used as provided by the sequencing facility.

Contamination screening on read level

To filter out reads possibly originating from contamination, a FastQ Screen like (FQS-like) approach was chosen. In brief, the reads are separated according to mapping behavior to different genomes. First, a database containing the genomes of *Daphnia magna* (Lee et al., 2019) and *D. pulex* (Ye et al., 2017) as positive controls and the human genome, the genome of the algae the sequenced individuals were fed on as well as several bacterial and viral genomes as negative controls was created. The database contains 108,163 sequences with a total length of 42.4 Gb (see Table M3.1). The accession numbers of the bacterial and viral genomes can be found in the corresponding lists deposited in Zenodo (doi:10.5281/zenodo.8069415).

Table M3.1. Database parts and corresponding sizes.

Species/group	Number of sequences	Total length (bp)
D. magna (GCA_001632505.1)	40,356	131,266,987
D. pulex (GCA_911175335.1)	1,822	156,418,198
Bacteria	18,448	37,434,584,693
Human (GCA_000001405.27)	594	3,257,319,537
Acutodesmus obliquus (GCA_002149895.1)	2,707	208,176,092
Virus	44,236	1,170,153,228

Illumina reads were mapped unpaired (forward and reverse reads separately) with NextGenMap 0.5.5 (Sedlazeck et al., 2013) and the options “--bam 1 --bin_size 4 --topn 1000”. PacBio subreads were mapped with minimap2 2.17-r941 (Li, 2018) and the options “-H -x map-pb”.

Custom scripts were used to filter reads and display the results (https://github.com/schellt/fqs-tools). The mapping results to the different database parts are displayed in Figure M3.1. Reads did only pass the filtering if they either did not map to the database at all or had at least one hit against one of the two *Daphnia* genomes. Paired Illumina reads were only returned if both reads pass the filtering. If only one read of a pair passed the filtering they were returned as unpaired. An overview of the effect of different read filtering steps on data volume is shown in Table M3.2.

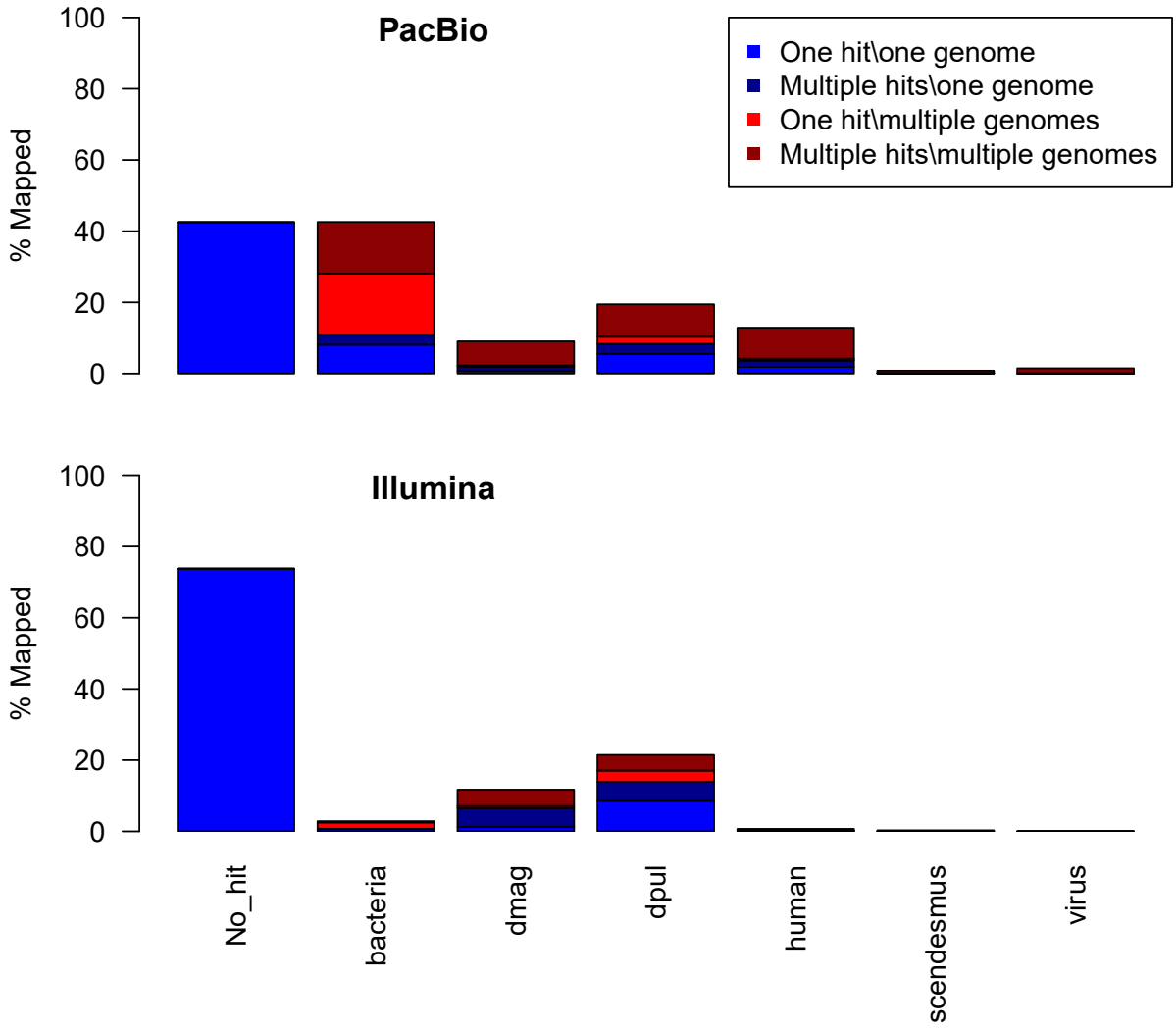


Figure M3.1. PacBio and Illumina reads mapped to the database to identify contamination.

Table M3.2. Comparison of data amount regarding the different preprocessing steps.

		Raw		Trimmed		FQS-like filtered	
		#reads	**Gb**	**#reads**	**Gb**	**#reads**	**Gb**
Illumina	paired	77,594,732	11.64	64,988,122	9.13	62,533,940	8.79
	unpaired	/	/	5,322,216	0.66	5,962,964	0.76
PacBio		/	/	1,679,290	11.52	1,121,227	6.84

RA assembly

All paired and unpaired contamination filtered Illumina reads as well as the contamination filtered PacBio reads were used as input for RA 0.2.1 (https://github.com/rvaser/ra).

Contamination screening on assembly level

To screen the resulting assembly, blobtools 1.0 (Laetsch & Blaxter, 2017) was used. The Illumina reads used as input for RA were mapped against the contigs using backmap.pl 0.1 (https://github.com/schellt/backmap), which integrates bwa mem 0.7.17-r1188 (Li, 2013), samtools 1.9-33-g2d34e15 (Li et al., 2009), Qualimap 2.2.1 (Okonechnikov et al., 2016), bedtools 2.26.0 (Quinlan & Hall, 2010) and R 3.5.0 (R Core Team, 2017). To assign Taxonomy IDs blastn 2.9.0+ (Camacho et al., 2009) was used to align the contigs against the complete nt database (-task megablast -outfmt '6 qseqid staxids bitscore' -max_target_seqs 1 -max_hsps 1 -evalue 1e-25). For plotting blobtools 1.1.1 was used.

Contamination with different bacteria can be clearly seen in Figure M3.2. In total 267 contigs with coverage below 10× and/or GC content above 50% were removed. The removed contigs add up to a total length of 22.97 Mb (see detailed distribution in Figure M3.3). To minimize false scaffolding later, PacBio reads mapping to the 267 contigs identified as contamination were removed. This resulted in 871,867 reads with a total length of 5.41 Gb which corresponds to 78% and 79% of the FQS-like filtered reads, respectively.

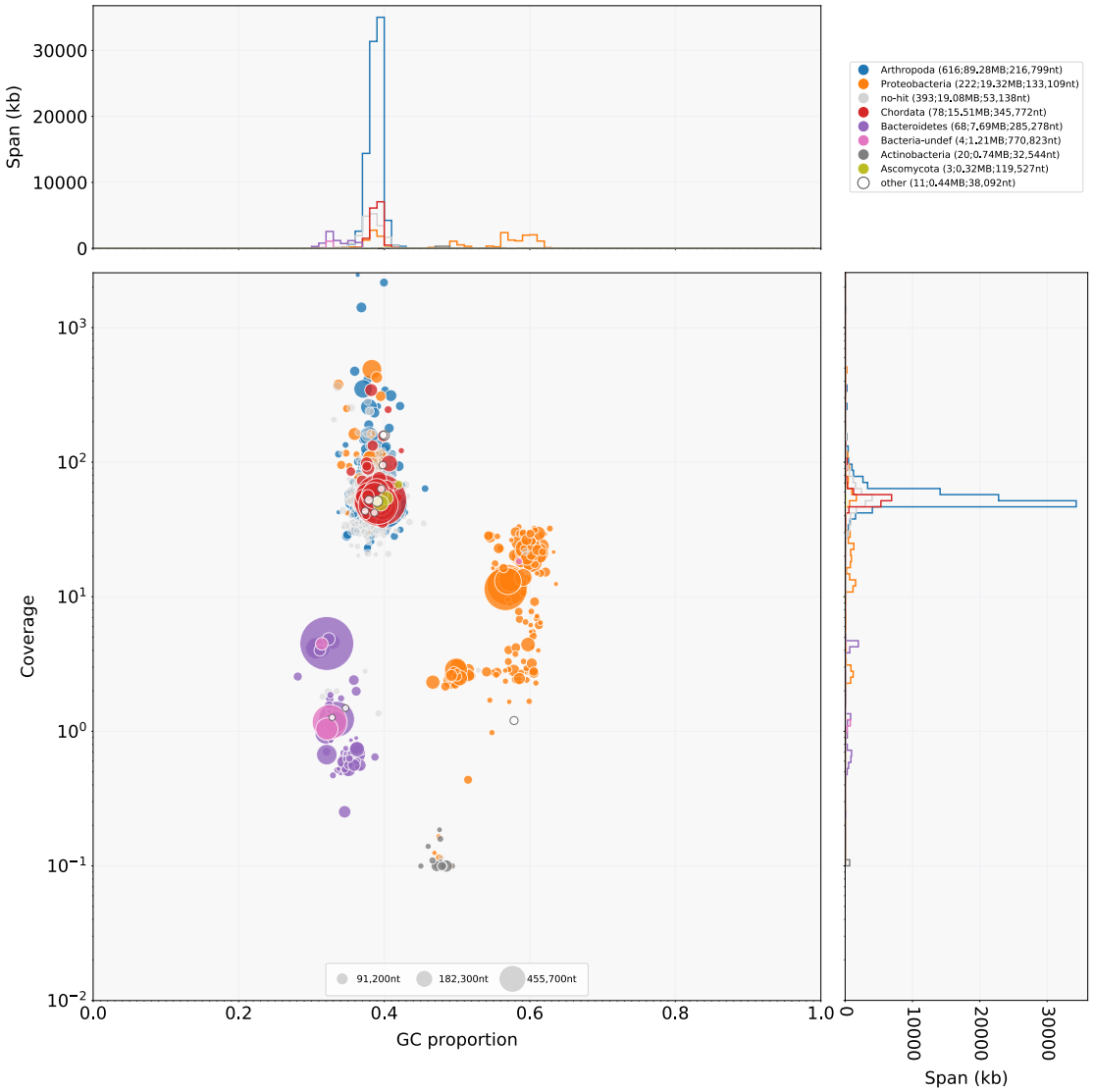


Figure M3.2. Blobplot of the initial RA assembly. Coverage is based on Illumina data only. Taxonomic assignment was conducted with blastn against the nt database.

Identifying and removing the mitochondrial genome

After assembly, the mitochondrial (mt) genome was searched with blastn 2.9.0+ by using the available *D. galeata* mt genomes (accession numbers LC177072.1, LC152879.1, LC177110.1, NC_034297.1, LC177071.1, LC177070.1) as queries against the contigs remaining after blobtools filtering. A total of 15 contigs had hits to at least one of these mt genomes. However, one single contig could be identified by calculating the ratio between alignment length and target sequence (contig) length. The mt genome could be distinguished by the maximum ratio 0.98 from random hits which reached maximally a ratio of 0.008. Finally, one single contig with a length of 15,636 bp was removed.

Scaffolding and gap closing

Scaffolding and gap closing was conducted in three iterations. Each iteration contains I) a scaffolding with SSPACE LongRead 1-1 (Boetzer & Pirovano, 2014), II) gap closing with LR_Gapcloser (https://github.com/CAFS-bioinformatics/LR_Gapcloser; commit 156381a) and III) polishing of the former gap parts (that were filled with PacBio reads only) with short reads using wtdbg2-racon-pilon.pl 0.4 (https://github.com/schellt/wtdbg2-racon-pilon), which used bwa mem 0.7.17-r1188, samtools 1.9, java 1.8.0_221 (Arnold et al., 2000) and Pilon 1.23 (Walker et al., 2014).

The blobtools filtered PacBio reads were used for scaffolding and gap closing. Coordinates of gaps produced in the scaffolding step were identified and written into bed format with the help of bedtools 2.28.0. Afterwards, the FQS-like filtered paired and unpaired Illumina reads were mapped via bwa mem (-a -c 10000 -t 60) to the gap closed scaffolds. Only reads aligning at least partially to former gap regions were saved to bam format as well as sorted and indexed with samtools. Finally, Pilon was run with additional options "---diploid ----threads 60" on the produced bam files. After running Pilon, the scaffold IDs were renamed to match the ones in the bed file again. Short read mapping and Pilon polishing was executed three times in iterative fashion using in all three polishing iterations the gap regions produced at the beginning from the long read scaffolding iteration. After the third iteration of scaffolding, gap closing and polishing the final scaffolds are yielded.

Assembly quality assessment

Contiguity was analyzed with Quast 5.0.2 (Gurevich et al., 2013) at different stages of the assembly process and its main results are represented in Table M3.3.

Table M3.3. Contiguity statistics of the single assembly steps.

	ra	ra-blobfilter-rmmt	ra-blobfilter-scaff1	ra-blobfilter-scaff2	ra-blobfilter-scaff3
#Sequences	1,415	1,147	473	370	346
Total length (Mb)	153.6	130.6	132.9	133.2	133.3
N50 (kb)	172.4	175.3	533.0	729.4	756.7
GC (%)	40.02	38.74	38.75	38.75	38.75
N's (%)	~0	~0	0.07	0.09	0.09

To look at mapping rate, coverage and insert size distribution, etc., backmap 0.3 (https://github.com/schellt/backmap) was used in combination with bwa mem 0.7.17-r1188, Minimap 2.17-r941, Samtools 1.9, Qualimap 2.2.1, bedtools 2.28.0, R 3.5.3, and MultiQC 1.8. Mapping the paired and unpaired FQS-like filtered Illumina reads and the blobtools filtered PacBio reads resulted in a mapping rate of 94.1% and 85.5%, respectively, according to Qualimap's bamqc. The insert size of the paired Illumina reads is narrowly distributed around

a median of 327 (Figure M3.3). The genome size was estimated based on mapped nucleotides and mode of the coverage distribution by backmap, resulting in 156.86 Mb and 178.03 Mb for Illumina (52×) and PacBio (26×) respectively. Additionally, the genome size estimated using a k-mer based approach with GenomeScope (150.6 Mb) can be found here: http://qb.cshl.edu/genomescope/analysis.php?code=WeYH4KGn4W7eTjIa1qhf.

To show absence of contamination in the assembly, a blastn search of the final scaffolds against the nt database was conducted as above, and the results plotted with blobtools 1.1.1. Completeness in terms of single copy core orthologs of the final scaffolds was assessed with BUSCO 3.0.2 (Simão et al., 2015), using the Arthropoda set (odb9) and the option ---long. This resulted in C: 95.7% [S: 94.7%, D: 1.0%], F: 0.8%, M: 3.5%, *N* = 1,066.

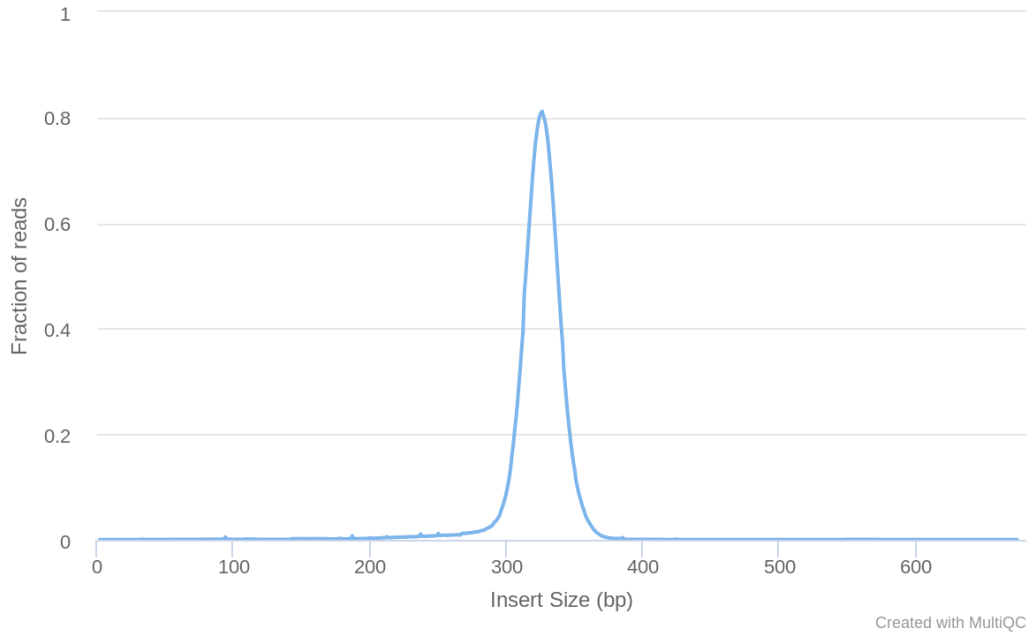


Figure M3.3. Insert size distribution of paired Illumina reads. Created with Qualimap and MultiQC.

Genome annotation

Before starting with the annotation, the fasta headers of the genome assembly and the *D. galeata* transcriptome assembly (HAFN01.1) were simplified with Augustus' simplifyFastaHeaders.pl.

Repeat library creation and repeat masking

To identify *D. galeata* specific repeats, RepeatModeler 2.0 (Smit & Hubley, 2008–2015) in combination with RepeatMasker 4.1.0 (Smit et al., 2013–2015) including RepBase release 20181026 (W. Bao et al., 2015), RECON 1.08 (Z. Bao & Eddy, 2002), RepeatScout 1.0.6 (Price et al., 2005), Tandem Repeats Finder 4.0.9 (Benson, 1999) and RMBlast 2.9.0+. RepeatModeler was run with the options "-pa 10 -LTRStruct" and resulted in 1,115 families

with a total length of ~1 Mb. Detailed distribution of repeat family classification can be seen in Table M3.4.

Table M3.4. Distribution of repeat family classification.

Classification	Number of families	Total length (bp)
Unknown	796	474,907
LTR	145	249,242
DNA	93	155,424
LINE	35	70,649
tRNA	31	43,674
SINE	13	17,071
RC	5	7,080
Simple repeat	4	364
SINE	2	5,442
snRNA	1	3,215

The yielded repeat families were combined with 237 *D. pulex* and 1 *D. pulicaria* repeat sequences from RepBase release 20181026 to create the final repeat library with 1,353 sequences and a total length of 1.78 Mb. The genome assembly was then soft masked for later usage within Augustus with the final repeat library. To do so, RepeatMasker 4.1.0 was applied with the search engine rmblastn 2.9.0+, Tandem Repeats Finder 4.0.9 and the additional options “-xsmall -no_is -e ncbi -pa 10 -s”. Afterwards, 21.9% of the assembly was masked. The distribution of masked fraction per repeat element can be found in Table M3.5.

Table M3.5. Repeat masking results as provided by RepeatMasker. Most repeats fragmented by insertions or deletions have been counted as one element.

Classification	Number of elements	Percentage of assembly
Unknown	48,249	10.18
LTR	7,824	5.25
DNA	7,145	2.29
Simple repeat	60,548	1.58
LINE	2,631	1.36
snRNA	1,591	0.49
Low complexity	12,696	0.43
SINE	849	0.19
RC	444	0.13

Creation of initial gene prediction models

Three different gene prediction models were produced with Augustus, GeneMark, and SNAP. The detailed procedure is described below.

One model was created using autoAug.pl from Augustus 3.3.2 (Stanke et al., 2008) in combination with PASA 2.4.1 (Haas et al., 2003) and GMAP 2019-09-12 (T. D. Wu & Watanabe, 2005). As Input for autoAug.pl the soft masked assembly and the *D. galeata*

transcriptome were used (HAFN01.1, Huylmans et al., 2016) as well as the additional options “-v --v --v ----pasa ----useGMAPforPASA -----noninteractive”.

A second model was created using GeneMark ET 4.48_3.60_lic (Lomsadze et al., 2005). First different RNAseq reads (ERR1551794 and ERR1551390–ERR1551397) were trimmed with autotrim.pl. The trimmed RNAseq reads were mapped paired and unpaired separately with HISAT 2.1.0 (Kim et al., 2019) and the additional option “-p 70”. The two resulting bam files were merged and sorted (--l 9 --@ 10) using samtools 1.9. Afterwards Augustus’ bam2hints with the additional options “----minintronlen=20 ----maxintronlen=500000” was used to produce a gff file containing possible introns. This gff file was filtered using filterIntronsFindStrand.pl (included in Augustus) with the additional option “----score”. Finally, gmes_petap.pl from GeneMark was run with the unmasked genome assembly, the filtered intron gff file and the additional options “----v ----cores=88 ----max_intron=400000”.

A third model was created with busco2snap.pl 0.1 (https://github.com/schellt/busco2snap) in combination with SNAP 2006-07-28 (Korf, 2004), Augustus 3.3.2, and blastp 2.9.0+. In brief, the intron-exon-boundaries of 1,010 complete and single copy BUSCOs are predicted with the Augustus model created with BUSCO. If more than one gene is predicted per locus a blastp search against the ancestral variants of the BUSCO genes is used to select the most likely one. Afterwards, the exon annotation is converted from gff to zff and finally the SNAP model is created with its tools fathom, forge and hmm-assembler.pl.

Structural annotation

The structural annotation was conducted in MAKER 2.31.10 (Holt & Yandell, 2011) in combination with blast 2.10.0+, RepeatMasker 4.1.0, Exonerate 2.4.0 (Slater & Birney, 2005), SNAP 2013-11-29, GeneMark 3.60_lic, Augustus 3.3.2, and tRNAscan-SE 1.3.1 (Lowe & Eddy, 1997). The following input sequences were used for MAKER: the unmasked genome assembly, the species' own transcriptome assembly (HAFN01.1) as ESTs, the complete Swiss-Prot 2019_10 (UniProt Consortium, 2019), and the protein sequences resulting from *D. magna* (Lee et al., 2019) as well as *D. pulex* (Ye et al., 2017) genome annotations as protein evidence. Repeat masking was conducted with the above described final repeat library for RepeatMasker and MAKER’s te_proteins.fasta for RepeatRunner. For model-based gene prediction, the above described models from SNAP, GeneMark, and Augustus were fed into MAKER. The options est2genome, protein2genome, trna, and alt_splice were switched on. The minimum protein length was set to 10aa. MAKER was compiled and executed with mpich 3.3.2 and the additional option “-fix_nucleotides”. With this, the first round of MAKER was completed.

Afterwards, one gff file with and one without the assembly sequence were created using MAKER's gff3_merge. From the gff without assembly sequence the tracks "est2genome", "protein2genome", and "repeatmasker"/"repeatrunner" were extracted and saved in three separate gff files. The gff file with assembly sequence was used as input for MAKER's maker2zff. The resulting files were processed with fathom (fathom genome.ann genome.dna -categorize 1000 && fathom -export 1000 -plus uni.ann uni.dna), forge (forge export.ann export.dna), and hmm-assembler.pl to create a new SNAP model based on the first round of MAKER.

The Augustus model was retrained with autoAug.pl as above, except that the gff file without assembly sequence from the first round's results was specified as training set. The second round of MAKER was conducted as the first round, but instead of specifying the same evidence and repeat sequences, the gff files containing the tracks est2genome, protein2genome and repeats as est_gff, protein_gff and rm_gff were used respectively. Furthermore, the retrained/new models from Augustus and SNAP were used.

After finishing the second round of MAKER, the Augustus and SNAP model were retrained/newly created as above. The third and final round of MAKER was run as the second one, except for updating the Augustus and SNAP model. The options est2genome and protein2genome were switched off, whereas the option always_complete was switched on. Finally, a gff file containing the maker track only without the assembly sequence was created with gff3_merge. The transcripts and protein fasta files of the annotated genes were created with fasta_merge.

Structural annotation quality assessment

The final gff file was processed with a custom script to count and calculate several contiguity statistics of the structural annotation except for tRNAscan results. Furthermore, the annotated protein set was analyzed using BUSCO 3.0.2 in combination with the Arthropoda set (odb9) and DOGMA 3.4 (Dohmen et al., 2016) in combination with Pfam scan 1.6 and DOGMA's reference set for Arthropoda. The results are compared in Table M3.6.

As further quality criterion the Annotation Editing Distance (AED) was compared between the three MAKER rounds and is visualized in Figure M3.4.

Table M3.6. Structural annotation statistics. Contiguity statistics of the annotation were calculated excluding tRNAscan results. BUSCO 3.0.2 was executed in protein mode for the different MAKER rounds. Conserved Domain Arrangements (CDAs) were searched with Pfam scan 1.6 and DOGMA 3.4.

	Assembly	Round 1	Round 2	Round 3
Number				
Gene		15,513	15,429	15,845
mRNA		16,471	16,275	16,774
Exon		121,995	119,033	117,364
CDS		130,242	123,265	119,402
Mean number of				
mRNAs/gene		1.06	1.05	1.06
Exons/mRNA		7.41	7.31	7.00
CDSs/mRNA		7.91	7.57	7.12
Median length (bp)				
Gene		2,145	2,187	2,097
mRNA		2,236	2,248	2,142
Exon		163	164	167
Intron		74	74	74
CDS		150	151	152
Total space (bp)				
Gene		54,765,381	52,605,643	51,689,473
mRNA		54,765,288	52,605,588	51,689,329
Exon		29,662,778	29,222,288	29,314,592
CDS		25,750,445	25,271,022	25,132,876
Single				
Exon mRNA		1,015	746	663
CDS mRNA		1,104	854	710
BUSCO (%) *N* = 1,066				
Complete	95.7	93.9	93.9	94.3
Single copy	94.7	91.0	91.0	91.7
Duplicated	1.0	2.9	2.9	2.6
Fragmented	0.8	0.8	1.0	0.7
Missing	3.5	5.3	5.1	5.0
CDAs (%) *N* = 4,222		93.94	93.89	93.63

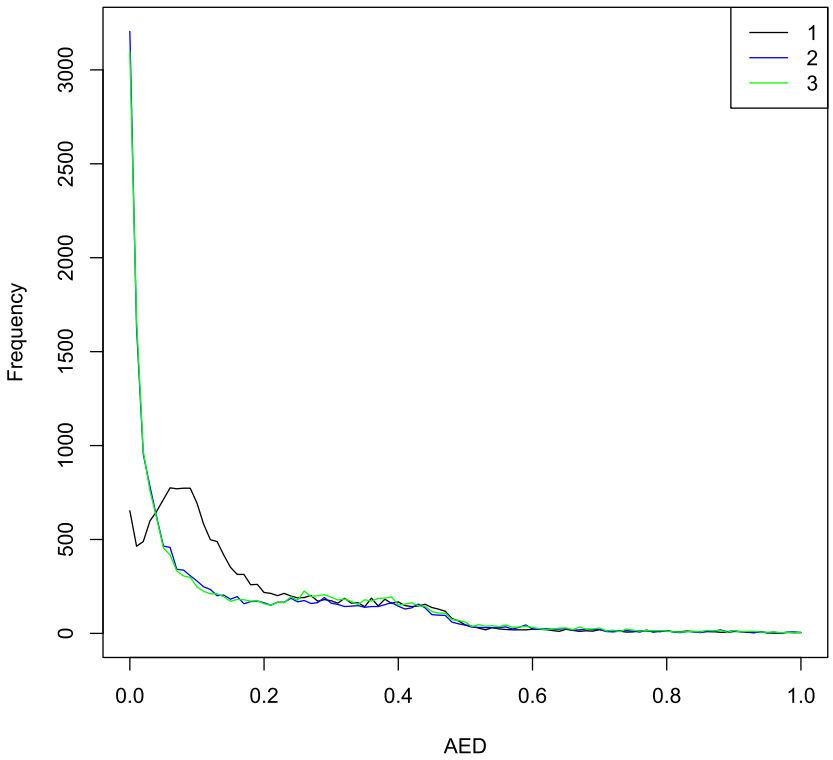


Figure M3.4. Annotation Editing Distance (AED) distribution of the different MAKER iterations. The AED shows congruency between the annotation and evidence alignments. Low values indicate high congruency.

Functional annotation

The functional annotation of the annotated protein set was conducted using InterProScan 5.39-77.0 (Jones et al., 2014) in combination with the databases/analyses CDD 3.17 (Lu et al., 2020), Coils 2.2.1 (Lupas et al., 1991), Gene3D 4.2.0 (Yeats et al., 2006), Hamap 2019_01 (Pedruzzi et al., 2015), MobiDBLite 2.0 (Necci et al., 2017), PANTHER 14.1 (Mi et al., 2013), Pfam 32.0 (El-Gebali et al., 2019), Phobius 1.01 (Käll et al., 2004), PIRSF 3.02 (C. H. Wu et al., 2004), PRINTS 42.0 (Attwood et al., 2012), ProSitePatterns 2019_01 and ProSiteProfiles 2019_01 (Sigrist et al., 2013), SFLD 4 (Akiva et al., 2014), SignalP_EUK 4.1 (Nielsen, 2017), SMART 7.1 (Letunic & Bork, 2018), SUPERFAMILY 1.75 (Wilson et al., 2009), TIGRFAM 15.0 (Haft et al., 2001), and TMHMM 2.0c (Krogh et al., 2001). The lookup for pathways and Gene Ontology (GO) terms was switched on.

To additionally assign putative names, an alignment of the annotated protein set against the Swiss-Prot 2019_10 was conducted using blastp 2.9.0+ with the options “-num_threads 70 -max_hsps 1 -max_target_seqs 1 -outfmt 6”. In total 15,898 (94.78%) and 15,960 (95.15%) protein sequences could be annotated by InterProScan and blast against Swiss-Prot, respectively. In combination, a functional annotation for 16,675 (99.41%) protein sequences could be achieved. A detailed overview of the functional annotated sequences per database or search algorithm is shown in Table M3.7.

Table M3.7. Functional annotation success of the different databases and search algorithms.

Database/analysis	Number of different protein sequences
Interpro	
CDD	5,550
Coils	3,191
Gene3D	9,977
Hamap	279
MobiDBLite	6,661
PANTHER	13,040
Pfam	11,567
Phobius	6,876
PIRSF	793
PRINTS	2,565
ProSitePatterns	3,169
ProSiteProfiles	6,378
SFLD	84
SignalP_EUK	2,951
SMART	5,444
SUPERFAMILY	9,816
TIGRFAM	721
TMHMM	4,211
GO	9,555
Reactome	3,927
Total	15,898
Total (%)	94.78
blast	
Swiss-Prot	15,960
Swiss-Prot (%)	95.15
Overall total	16,737
Overall total (%)	99.78

The results of the functional annotation were then included in the gff and the transcript/protein fasta files using adapted versions of MAKER's maker_functional_gff and maker_functional_fasta, respectively (https://github.com/schellt/maker-functional).

ANGSD

To confirm the result of the GATK based variant calling, we estimated genotype likelihoods using ANGSD v0.931-2-gfd2a527 with the following parameters: angsd -P 32 -bam $BAM -ref $REF -out $OUT -GL 2 -doGlf 2 -doMajorMinor 1 -doMaf 1 -SNP_pval 1e-6 -uniqueOnly 1 -remove_bads 1 -only_proper_pairs 1 -trim 0 -C 50 -baq 1 -minMapQ 20 -minQ 20 -setMinDepth 400 -setMaxDepth 2000 -minMaf 0.05 -doCounts 1, and performed admixture analysis using NGSadmix and PCA analysis using PCAngsd v1.02.

Supplementary materials and methods references

Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., Custer, A. F., Hicks, M. A., Huang, C. C., Lauck, F., Mashiyama, S. T., Meng, E. C., Mischel, D., Morris, J. H., Ojha, S., Schnoes, A. M., Stryke, D., Yunes, J. M., Ferrin, T. E., Holliday, G. L., & Babbitt, P. C. (2014). The Structure-Function Linkage Database. *Nucleic Acids Research*, *42*(D1), D521–D530. https://doi.org/10.1093/nar/gkt1130

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnold, K., Gosling, J., & Holmes, D. (2000). *The Java Programming Language, Third Edition*. Addison-Wesley.

Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., Romá-Mateo, C., Theodosiou, A., & Mitchell, A. L. (2012). The PRINTS database: A fine-grained protein sequence annotation and analysis resource—its status in 2012. *Database: The Journal of Biological Databases and Curation*, *2012*, bas019. https://doi.org/10.1093/database/bas019

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 11. https://doi.org/10.1186/s13100-015-0041-9

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, *12*(8), 1269–1276. https://doi.org/10.1101/gr.88502

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580. https://doi.org/10.1093/nar/27.2.573

Boetzer, M., & Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, *15*(1), 211. https://doi.org/10.1186/1471-2105-15-211

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Dohmen, E., Kremer, L. P. M., Bornberg-Bauer, E., & Kemena, C. (2016). DOGMA: Domain-based transcriptome and proteome quality assessment. *Bioinformatics*, *32*(17), 2577–2581. https://doi.org/10.1093/bioinformatics/btw231

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., & White, O. (2001). TIGRFAMs: A protein family resource for the functional identification of proteins. *Nucleic Acids Research*, *29*(1), 41–43. https://doi.org/10.1093/nar/29.1.41

Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*(1), 491. https://doi.org/10.1186/1471-2105-12-491

Huylmans, A. K., López Ezquerra, A., Parsch, J., & Cordellier, M. (2016). De Novo Transcriptome Assembly and Sex-Biased Gene Expression in the Cyclical Parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution*, *8*(10), 3120–3139. https://doi.org/10.1093/gbe/evw221

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, *338*(5), 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*(1), 59. https://doi.org/10.1186/1471-2105-5-59

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Lee, B.-Y., Choi, B.-S., Kim, M.-S., Park, J. C., Jeong, C.-B., Han, J., & Lee, J.-S. (2019). The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquatic Toxicology*, *210*, 69–84. https://doi.org/10.1016/j.aquatox.2019.02.009

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, *46*(D1), D493–D496. https://doi.org/10.1093/nar/gkx922

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., & Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Research*, *33*(20), 6494–6506. https://doi.org/10.1093/nar/gki937

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964. https://doi.org/10.1093/nar/25.5.955

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research*, *48*(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science*, *252*(5009), 1162–1164. https://doi.org/10.1126/science.252.5009.1162

Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, *41*(D1), D377–D386. https://doi.org/10.1093/nar/gks1118

Necci, M., Piovesan, D., Dosztányi, Z., & Tosatto, S. C. E. (2017). MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics*, *33*(9), 1402–1404. https://doi.org/10.1093/bioinformatics/btx015

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. In D. Kihara (Ed.), *Protein Function Prediction: Methods and Protocols* (pp. 59–73). Springer. https://doi.org/10.1007/978-1-4939-7015-5_6

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., De Castro, E., Baratin, D., Cuche, B. A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I., & Bridge, A. (2015). HAMAP in 2015: Updates to the protein family classification and annotation system. *Nucleic Acids Research*, *43*(D1), D1064–D1070. https://doi.org/10.1093/nar/gku1002

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, *21*(suppl_1), i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

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

Sedlazeck, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, *29*(21), 2790–2791. https://doi.org/10.1093/bioinformatics/btt468

Sigrist, C. J. A., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, *41*(D1), D344–D347. https://doi.org/10.1093/nar/gks1067

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), 31. https://doi.org/10.1186/1471-2105-6-31

Smit, A. F. A., & Hubley, R. (2008–2015). *RepeatModeler Open-1.0.* Accessed November 2021. https://www.repeatmasker.org/RepeatModeler/

Smit, A. F. A., Hubley, R., & Green, P. (2013–2015). *RepeatMasker Open-4.0.* Accessed November 2021. https://www.repeatmasker.org/RepeatMasker/

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Waldvogel, A.-M., Wieser, A., Schell, T., Patel, S., Schmidt, H., Hankeln, T., Feldmeyer, B., & Pfenninger, M. (2018). The genomic footprint of climate adaptation in *Chironomus riparius*. *Molecular Ecology*, *27*(6), 1439–1456. https://doi.org/10.1111/mec.14543

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., & Gough, J. (2009). SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Research*, *37*(suppl_1), D380–D386. https://doi.org/10.1093/nar/gkn762

Wu, C. H., Nikolskaya, A., Huang, H., Yeh, L. L., Natale, D. A., Vinayaka, C. R., Hu, Z., Mazumder, R., Kumar, S., Kourtesis, P., Ledley, R. S., Suzek, B. E., Arminski, L., Chen, Y., Zhang, J., Cardenas, J. L., Chung, S., Castro-Alvear, J., Dinkov, G., & Barker, W. C. (2004). PIRSF: Family classification system at the Protein Information Resource. *Nucleic Acids Research*, *32*(suppl_1), D112–D114. https://doi.org/10.1093/nar/gkh097

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Ye, Z., Xu, S., Spitze, K., Asselman, J., Jiang, X., Ackerman, M. S., Lopez, J., Harker, B., Raborn, R. T., Thomas, W. K., Ramsdell, J., Pfrender, M. E., & Lynch, M. (2017). A New Reference Genome Assembly for the Microcrustacean *Daphnia pulex*. *G3: Genes, Genomes, Genetics*, *7*(5), 1405–1416. https://doi.org/10.1534/g3.116.038638

Yeats, C., Maibaum, M., Marsden, R., Dibley, M., Lee, D., Addou, S., & Orengo, C. A. (2006). Gene3D: Modelling protein structure, function and evolution. *Nucleic Acids Research*, *34*(suppl_1), D281–D284. https://doi.org/10.1093/nar/gkj057

Supplementary figures

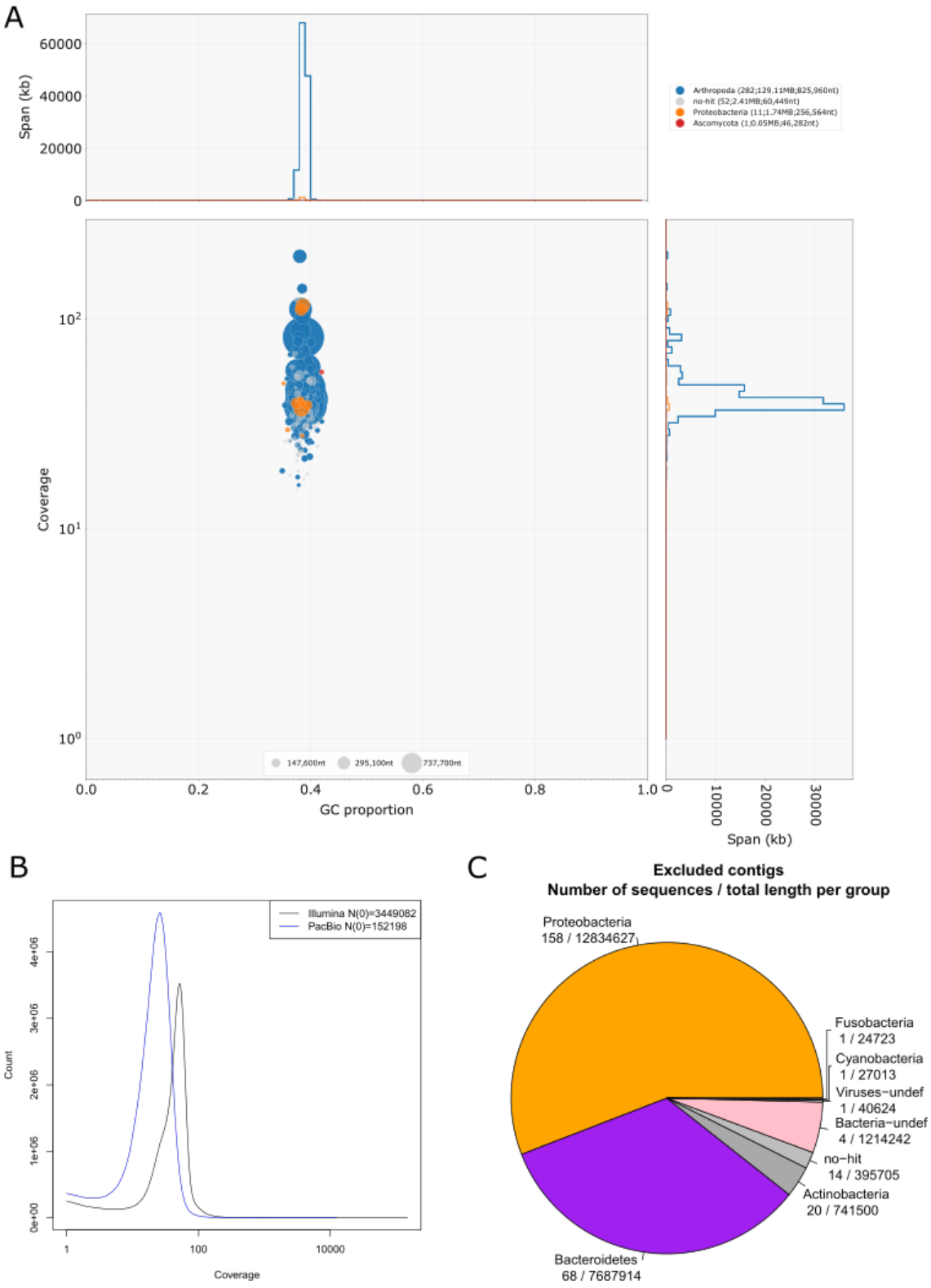


Figure S3.1. Assembly quality assessment. (A) Blobplot of the final assembly. Coverage is based on Illumina and PacBio data. Taxonomic assignment was conducted with blastn against the nt database. (B) Coverage distribution per sequencing technology of the final assembly (C) Excluded contigs. The pie chart displays the total length of the excluded contigs per taxonomic group. Below the group name the number of contigs and their total length is specified.

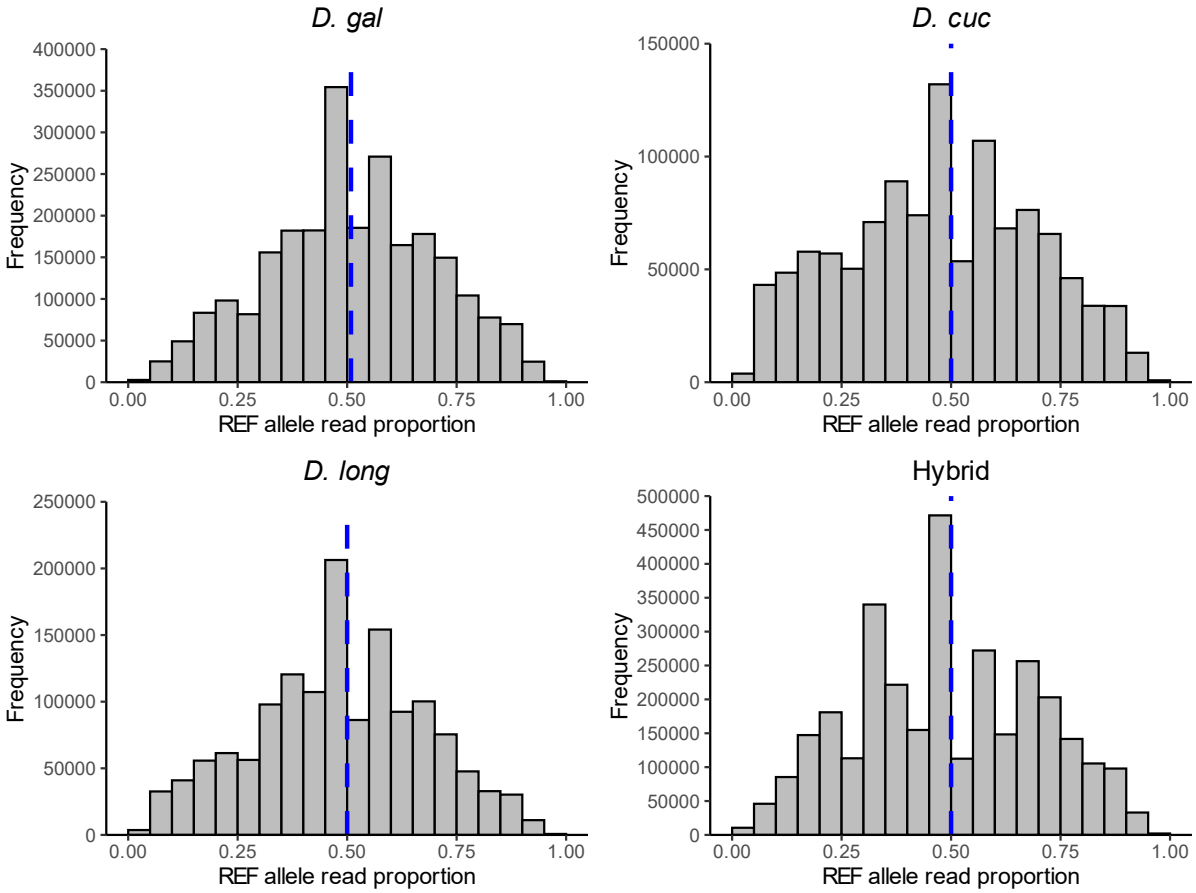


Figure S3.2. Distribution of reference and alternative alleles observed at heterozygous genotypes for all individuals which were unambiguously assigned to either of the parental species clusters *D. galeata*, *D. cucullata*, or *D. longispina* or classified as hybrids. The centering of this distribution around the blue median line suggests that there is no biased mapping of the reference alleles.

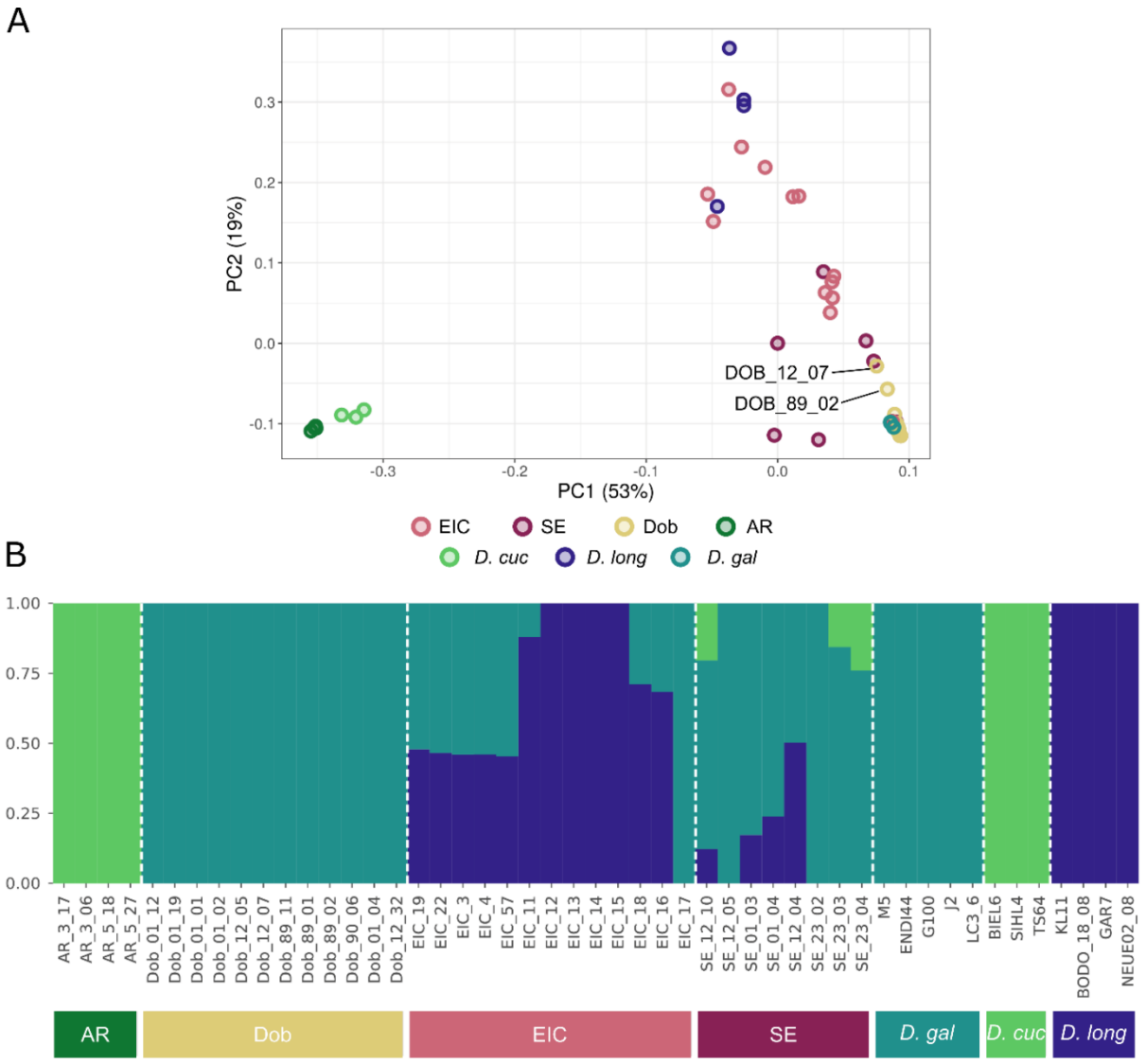


Figure S3.3. (A) PCA plot obtained with PCAngsd. (B) Admixture plot obtained with $K = 3$ in NgsAdmix. *D. gal: D. galeata*, *D. long*: *D. longispina*, *D. cuc*; *D. cucullata*. Bottom bars are color coded to match the color scheme used in panel A.

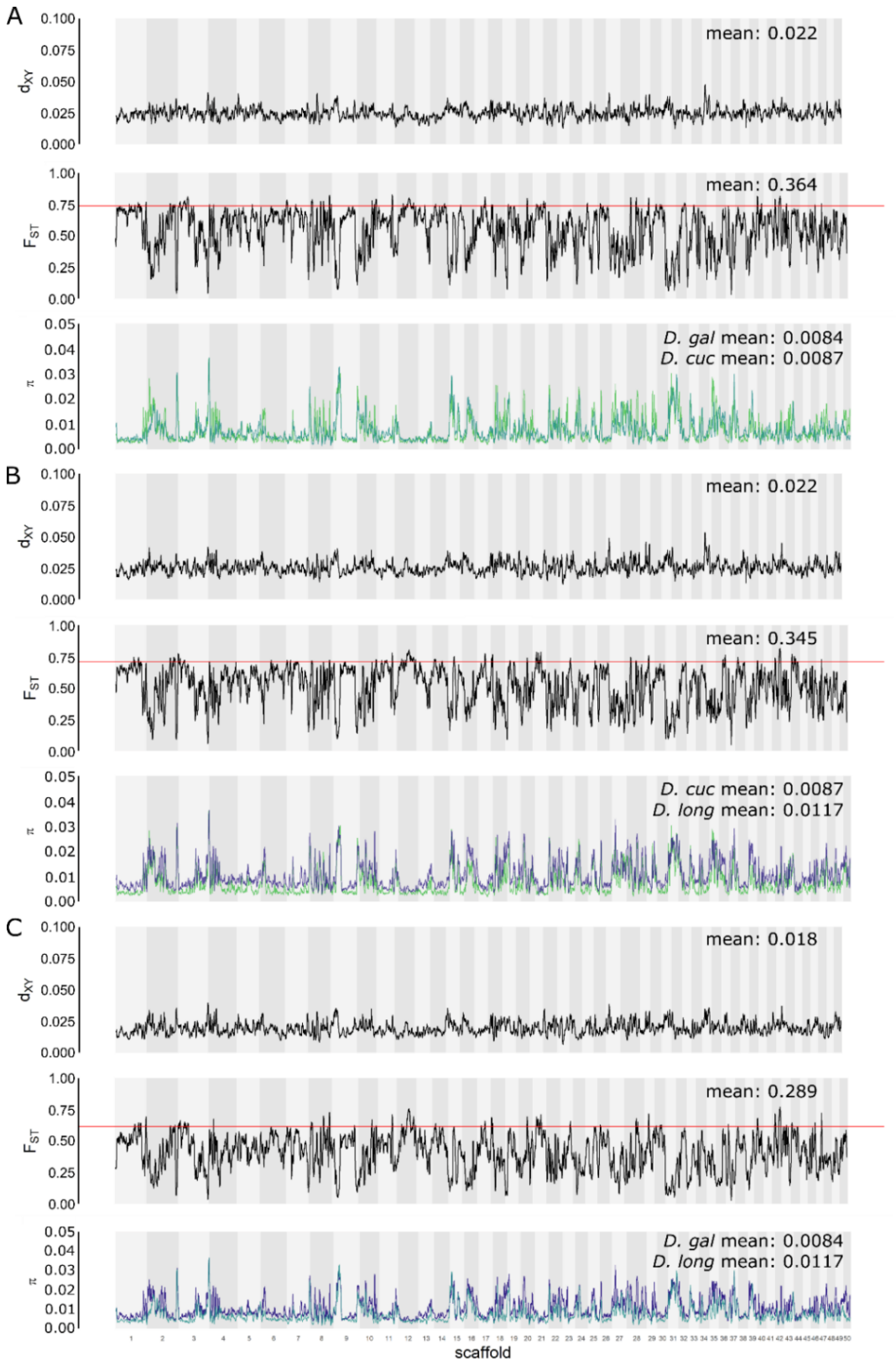


Figure S3.4. Window-based statistics reduced to one individual per population for the pairs (A) *D. galeata*/*D. cucullata,* (B) *D cucullata*/*D. longispina* and (C) *D. galeata*/*D. longispina*, shown for the 50 largest scaffolds in 100-kb windows with 10-kb step size – calculations are for one randomly chosen individual representing each population per species. In each panel from top to bottom: d_{xy} values, pairwise F_{ST} values with a red horizontal line indicating the 95th percentile, nucleotide diversity (π) for *D. galeata* (teal), *D. longispina* (dark blue) and *D. cucullata* (lime green).

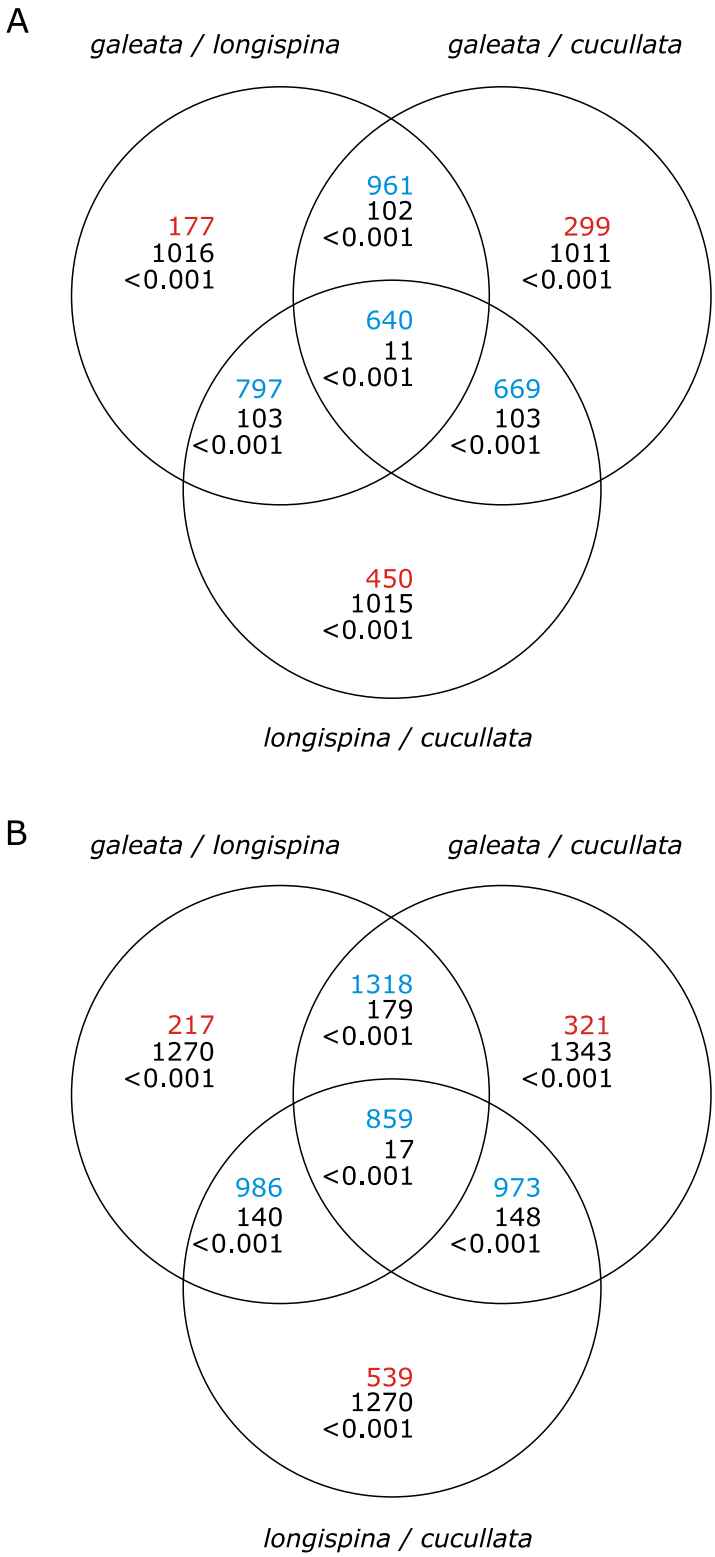


Figure S3.5. Venn diagram for (A) outlier windows (based on F_{ST} values) (B) genes within these outlier windows. The upper number in each region of the diagram is the observed number of items in the region, the middle number is the mean expected number according to 1000 randomizations and the lower number gives the probability of this outcome. The observed number is red when smaller than expected and blue when larger. *galeata* / *longispina*: set of elements found in the comparison between *D. galeata* and *D. longispina* in Figure 3.3C. *galeata* / *cucullata*: set of elements found in the comparison between *D. galeata* and *D. cucullata* in Figure 3.3A. *longispina* / *cucullata*: set of elements found in the comparison between *D. longispina* and *D. cucullata* in Figure 3.3B.

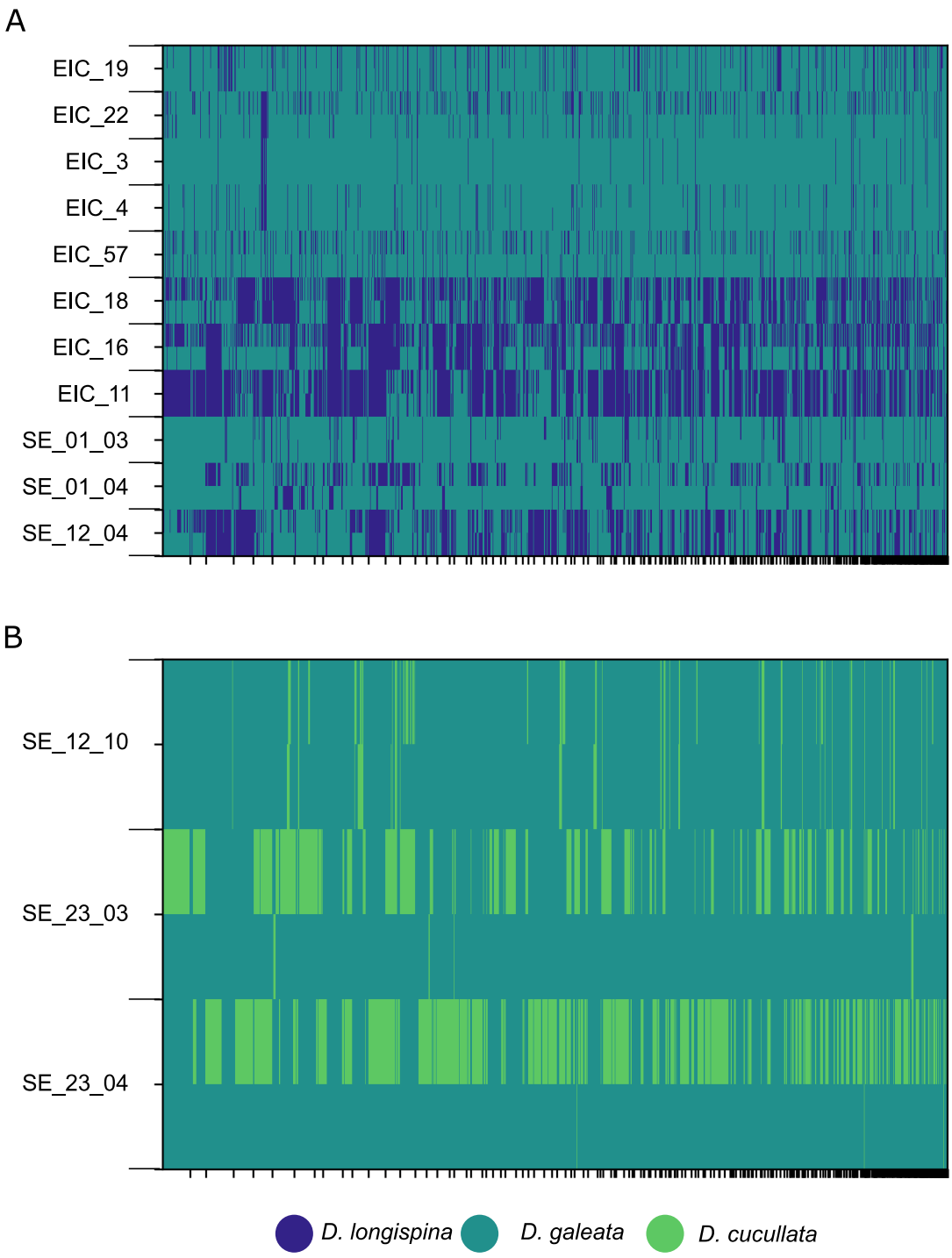


Figure S3.6. Local ancestry inference for: (A) *D. galeata*×*D. longispina* individuals. (B) 3 *D. galeata*×*D. cucullata* individuals. Ancestry tracts are visualized on the y-axis in two rows for the two haplotypes of one individual and on the x-axis along the genomic scaffolds.

Supplementary tables

Due to their size, Table S3.2 and S3.6 can be found in Zenodo (doi:10.5281/zenodo.8069415) and the complete supplementary material is online at Genome Biology and Evolution (https://academic.oup.com/gbe/article/13/12/evab267/6448229).

Table S3.1. Comparison of *Daphnia galeata* genome assembly with 4 other genome assemblies for *Daphnia pulex* and *Daphnia magna*. Overview obtained with Quast.

Assembly accession	**dgal_ra_scaffold_fasta**	**GCA_000187875.1**	**GCA_900092285.2**	**GCA_003990815.1**	**GCA_001632505.1**
Species	*Daphnia galeata*	*Daphnia pulex*	*Daphnia pulex*	*Daphnia magna*	*Daphnia magna*
Strain	M5	TCO	PA42	SK	Xinb3
# scaffolds	346	5,186	493	4,192	13,180
# scaffolds (≥ 0 bp)	346	5,186	493	4,192	28,801
# scaffolds (≥ 1000 bp)	346	5,186	493	4,192	6,576
Largest scaffold	2,950,711	4,193,030	7,584,612	16,359,456	3,718,170
Total length	133,304,630	197,206,209	189,550,516	122,937,721	124,519,356
Total length (≥ 0 bp)	133,304,630	197,206,209	189,550,516	122,937,721	129,543,483
Total length (≥ 1000 bp)	133,304,630	197,206,209	189,550,516	122,937,721	119,957,162
N50	756,671	642,089	1,160,003	10,124,675	417,686
N75	386,650	136,924	532,199	8,305,553	111,610
L50	48	75	36	5	56
L75	112	232	95	9	196
GC (%)	38.75	40.76	40.39	40.54	40.67
# *N*'s	120,845	38,598,801	4,006,006	8,297,703	23,479,941
# *N*'s per 100 kbp	90.65	19,573	2113.42	6,749.52	18,856

Table S3.2. Overview of the individuals included in the study – Sampled as: genotypes sampled as “egg” originate from ephippia isolated from sediment cores and were hatched in the lab or genotyped directly. Origin: culture – this genotype was maintained in the lab and raised in bulk before DNA extraction; WGA egg – DNA was prepared through Whole Genome Amplification on single resting eggs. Numbers of genotypes correspond to those with a calculated coverage >4. When lowering this threshold more samples from Dobersdorfer See and Eichbaumsee could be analyzed. Species categories are given for all included samples used for ancestry paintings and to calculate genome-wide nucleotide diversity (π), between-taxon differentiation (F_{ST}), and between-taxon divergence (d_{xy}).

Table S3.3. Sequencing results for resequencing samples, including number of raw reads, number of reads after trimming, mapping success, and mean read depth.

Sample	Location	Number of raw reads	Number of trimmed reads	Number of mapped reads	Mapped reads %	Mean read depth
AR_3_06	AR	21,288,352	19,870,634	16,218,682	81.62%	10.76
AR_3_17	AR	17,310,780	16,083,422	15,355,123	95.47%	10.68
AR_5_18	AR	17,520,952	16,408,352	15,546,068	94.74%	11.04
AR_5_27	AR	16,284,616	15,218,916	13,752,333	90.36%	9.42
SE_12_04	SE	42,684,512	41,338,814	13,098,094	31.68%	4.20
SE_01_03	SE	27,737,288	22,465,594	21,555,004	95.95%	12.79
SE_01_04	SE	23,550,114	21,910,332	21,299,188	97.21%	13.51
SE_23_02	SE	19,136,852	18,015,214	17,534,351	97.33%	12.06
SE_23_03	SE	19,537,960	18,398,560	17,893,952	97.26%	12.57
SE_23_04	SE	27,909,232	26,384,906	24,918,372	94.44%	17.26
SE_12_05	SE	20,708,746	19,447,996	19,180,692	98.63%	10.11
SE_12_10	SE	22,383,236	21,046,994	20,457,588	97.20%	10.68
Dob_01_12	Dob	15,720,858	13,187,394	12,954,419	98.23%	7.57
Dob_01_19	Dob	18,578,040	13,519,581	12,831,106	94.91%	7.17
Dob_01_01	Dob	23,835,814	20,372,380	20,076,825	98.55%	3.98
Dob_01_02	Dob	16,079,542	14,735,513	14,515,439	98.51%	8.95
Dob_12_05	Dob	26,637,800	19,573,870	19,304,659	98.62%	7.18
Dob_12_07	Dob	16,989,602	12,226,574	10,931,416	89.41%	5.53
Dob_89_11	Dob	17,954,342	13,776,236	13,531,141	98.22%	7.59
Dob_89_01	Dob	14,993,632	9,451,949	9,081,200	96.08%	4.42
Dob_89_02	Dob	14,248,786	10,049,553	9,106,535	90.62%	4.91
Dob_90_06	Dob	17,001,496	8,190,873	7,784,928	95.04%	2.62
Dob_01_04	Dob	15,012,108	10,526,529	10,331,250	98.14%	5.71
Dob_12_32	Dob	15,937,648	13,931,526	8,267,592	59.34%	4.94
EIC_19	EIC	13,190,044	4,548,433	3,808,405	83.73%	0.91
EIC_22	EIC	14,555,464	7,236,092	6,600,284	91.21%	3.12
EIC_3	EIC	29,058,912	11,454,006	5,260,372	45.93%	0.37
EIC_4	EIC	13,564,520	4,253,787	3,771,647	88.67%	0.85

EIC_57	EIC	14,239,924	6,395,056	5,616,724	87.83%	2.26
EIC_13	EIC	14,463,272	12,365,944	12,086,534	97.74%	6.46
EIC_14	EIC	16,917,644	12,015,092	11,760,413	97.88%	6.58
EIC_15	EIC	15,531,160	13,277,223	12,990,753	97.84%	6.56
EIC_18	EIC	14,952,714	10,033,233	9,367,630	93.37%	4.78
EIC_16	EIC	15,450,664	9,329,878	8,887,073	95.25%	4.58
EIC_17	EIC	15,392,966	12,502,445	12,080,674	96.63%	8.38
EIC_11	EIC	16,210,646	8,768,403	8,510,659	97.06%	3.25
EIC_12	EIC	19,121,834	2,788,735	2,594,290	93.03%	0.34
J2	J	79,048,948	73,020,526	65,159,914	89.24%	52.30
LC3_6	LC	90,918,942	85,065,214	61,806,326	72.66%	47.70
M5	M	77,594,732	64,988,122	54,986,148	84.61%	48.35
G100	G	39,015,978	31,495,250	29,785,801	94.57%	12.84
ENDI44	ENDI	41,567,550	39,657,286	37,973,867	95.76%	16.07
SIHL4	SIHL	55,802,822	43,669,966	40,870,371	93.59%	7.75
BIEL6	BIEL	44,908,864	42,806,350	39,963,285	93.36%	7.80
TS64	TS	95,251,354	77,115,502	71,785,929	93.09%	13.16
NEUE02_08	NEUE	64,689,492	54,264,594	51,069,789	94.11%	11.12
GAR7	GAR	66,935,478	52,166,782	48,728,229	93.41%	8.88
BODO_18_08	BODO	64,818,020	62,076,398	58,616,550	94.43%	21.81
KL11	KL	30,344,922	12,721,986	6,745,144	53.02%	0.74

Table S3.4. Limnological data of sampled lakes.

Location	Country	Coordinates	Max depth (m)	Trophy level	Stratification
Arendsee	Germany	52° 53′ 21″ N, 11° 28′ 27″ E	48.7	eutrophic	dimictic
Selenter See	Germany	4° 18′ 20″ N, 10° 26′ 43″ E	36.78	mesotrophic	dimictic
Eichbaumsee	Germany	53° 29′ 6″ N, 10° 6′ 11″ E	16	eutrophic	unknown
Dobersdorfer See	Germany	54° 19′ 7″ N, 10° 18′ 22″ E	18.8	eutrophic	polymictic

Table S3.5. Mitochondrial genome assembly summary.

Sample	mt genome length	Assembly program	Mitotype assignment
AR_3_06	15,884	GetOrganelle	cuc
AR_3_17	18,166	GetOrganelle	gal
AR_5_18	18,166	GetOrganelle	gal
AR_5_27	15,328	MitoZ	cuc
SE_12_04	15,411	MitoZ	gal
SE_01_03	15,308	MitoZ	gal
SE_01_04	15,977	GetOrganelle	gal
SE_23_02	14,915	MitoZ	gal
SE_23_03	15,327	MitoZ	cuc
SE_23_04	15,327	MitoZ	cuc
SE_12_05	15,389	MitoZ	gal
SE_12_10	14,929	MitoZ	cuc
Dob_01_12	15,711	MITObim	gal
Dob_01_19	14,980	MitoZ	gal
Dob_01_01	15,980	GetOrganelle	gal
Dob_01_02	15,257	MitoZ	gal
Dob_12_05	15,351	MitoZ	gal
Dob_12_07	15,271	MitoZ	gal
Dob_89_11	15,447	MitoZ	gal
Dob_89_01	15,271	MitoZ	gal
Dob_89_02	15,271	MitoZ	gal
Dob_90_06	15,739	MITObim	gal
Dob_01_04	16,158	GetOrganelle	gal
Dob_12_32	15,768	MITObim	gal
EIC_19	15,631	MitoZ	gal
EIC_22	15,294	MitoZ	gal
EIC_4	15,662	MITObim	gal
EIC_57	16,052	MITObim	gal
EIC_14	14,899	MitoZ	long
EIC_15	15,804	MITObim	gal
EIC_18	15,306	MitoZ	gal
EIC_16	14,914	MitoZ	gal
EIC_17	15,798	GetOrganelle	gal
J2	15,306	MitoZ	gal
LC3_6	15,975	GetOrganelle	gal

M5	15,808	GetOrganelle	gal
G100	15,978	GetOrganelle	gal
ENDI44	15,258	MitoZ	gal
SIHL4	16,432	GetOrganelle	cuc
BIEL6	16,251	GetOrganelle	cuc
TS64	14,979	MitoZ	cuc
NEUE02_08	16,077	GetOrganelle	long
GAR7	16,059	GetOrganelle	long
BODO_18_08	14,968	MitoZ	long

Table S3.6. Summary table of the protein-coding (PCG) and ribosomal RNA (rRNA) gene matrix.

Table S3.7. Data derived from ancestry painting analysis and based on the 95,302 fixed sites inferred from analyzing parental species genotypes, for EIC and DOB genotypes inferred as non-admixed. Maternal species attribution is based on mitochondrial phylogeny, hybrid attribution is based on the ADMIXTURE plot.

Sample	Hybrid Index	Heterozygosity	Maternal species	Admixture assignment
Dob_12_07	0.993	0.000	gal	gal
Dob_89_02	0.994	0.001	gal	gal
EIC_13	0.125	0.047	gal	long
EIC_14	0.860	0.013	long	long
EIC_15	0.199	0.083	gal	long
EIC_17	0.993	0.005	gal	gal
EIC_12	0.150	0.019	gal	long

Supplementary Information on Chapter 4

Supplementary figures

Due to their size, Figure S4.6 and S4.7 can be found in Zenodo (doi:10.5281/zenodo.8069314).

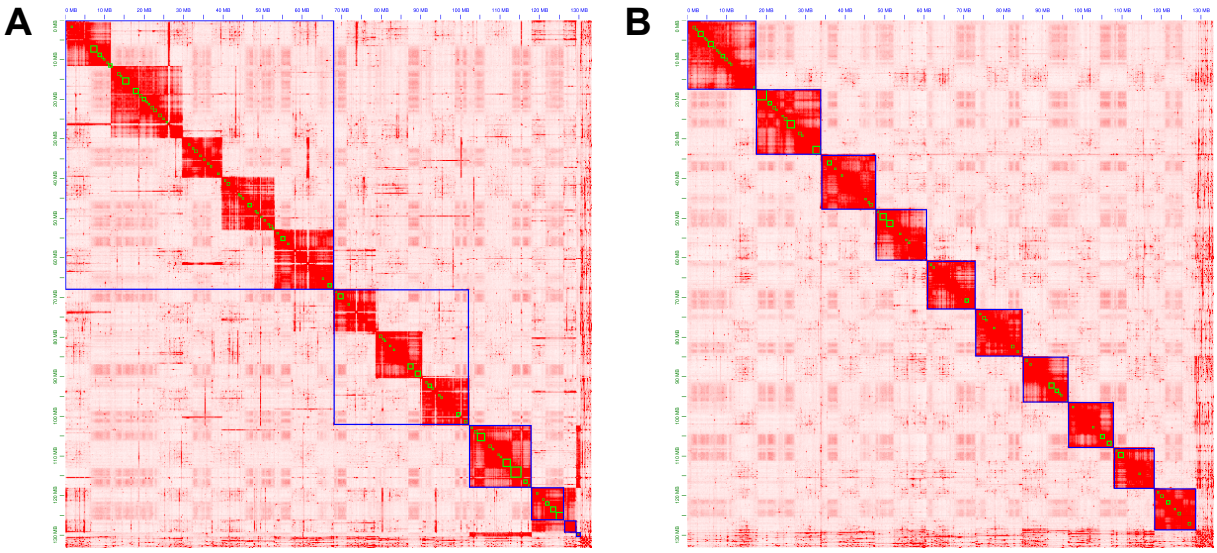


Figure S4.1. Hi-C contact map of the assembled *D. galeata* scaffolds (A) before and (B) after manual curation. Chromosome-scale scaffolds are shown in the blue rectangles.

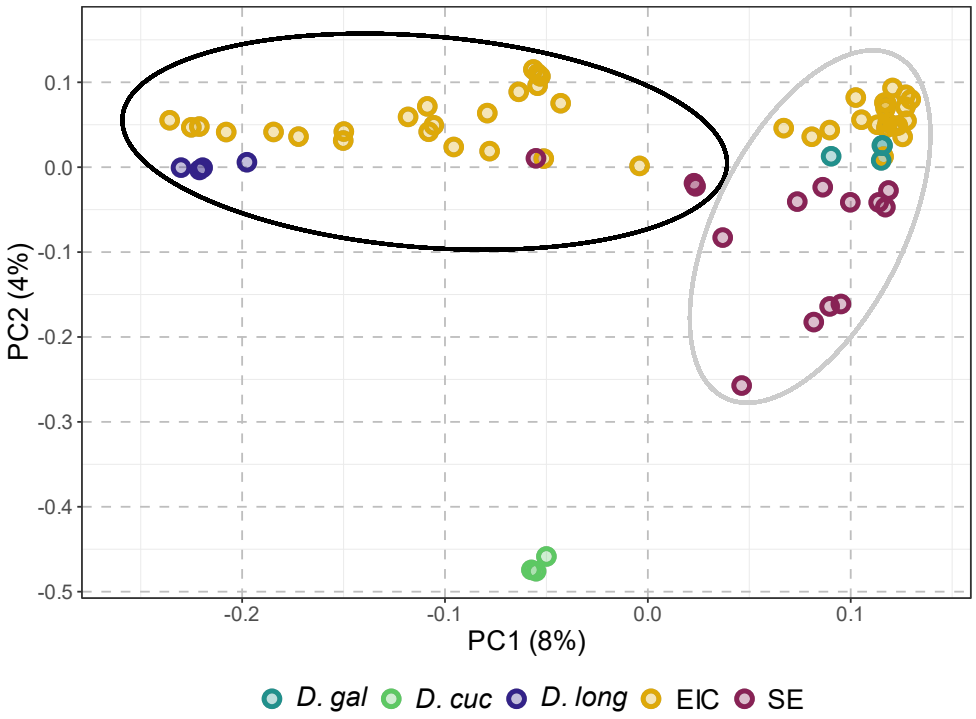


Figure S4.2. Genomic principal component analysis based on 607,440 SNPs. The samples from populations EIC and SE assigned as species pair *D. galeata* × *D. longispina* (black) or *D. galeata* × *D. cucullata* (gray) for the junctions analysis are circled.

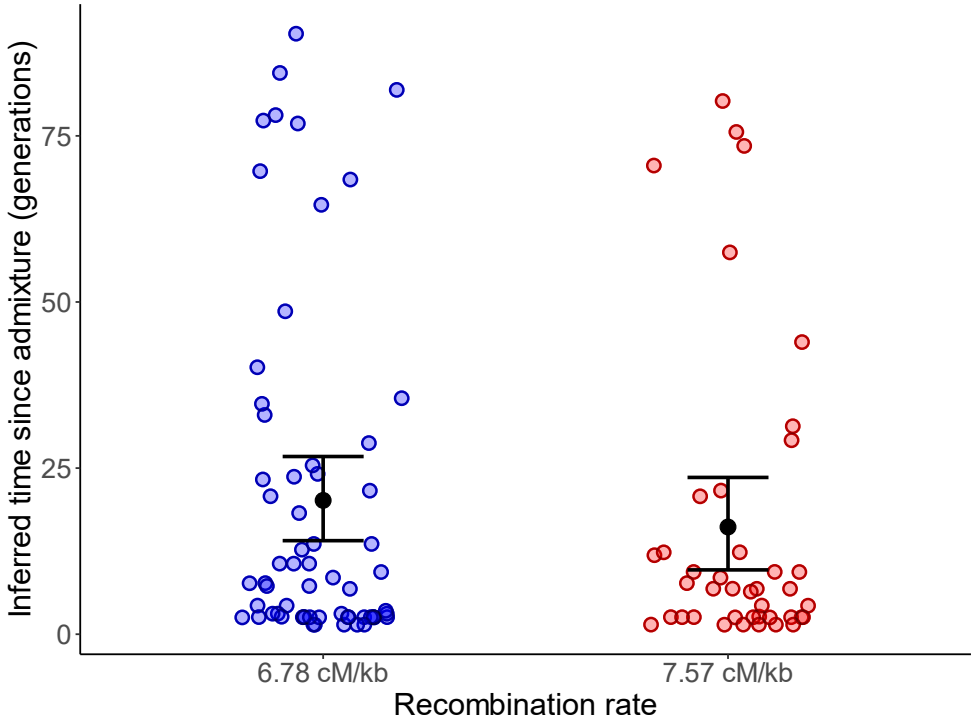


Figure S4.3. Inferred time since admixture for all 52 individuals using two different rates, 6.78 cM/kb from Dukić et al. (2016) and 7.57 cM/kb from Wersebe et al. (2023). The black dots indicate the bootstrapped mean estimated age, and the error bars indicate the corresponding 95% confidence interval.

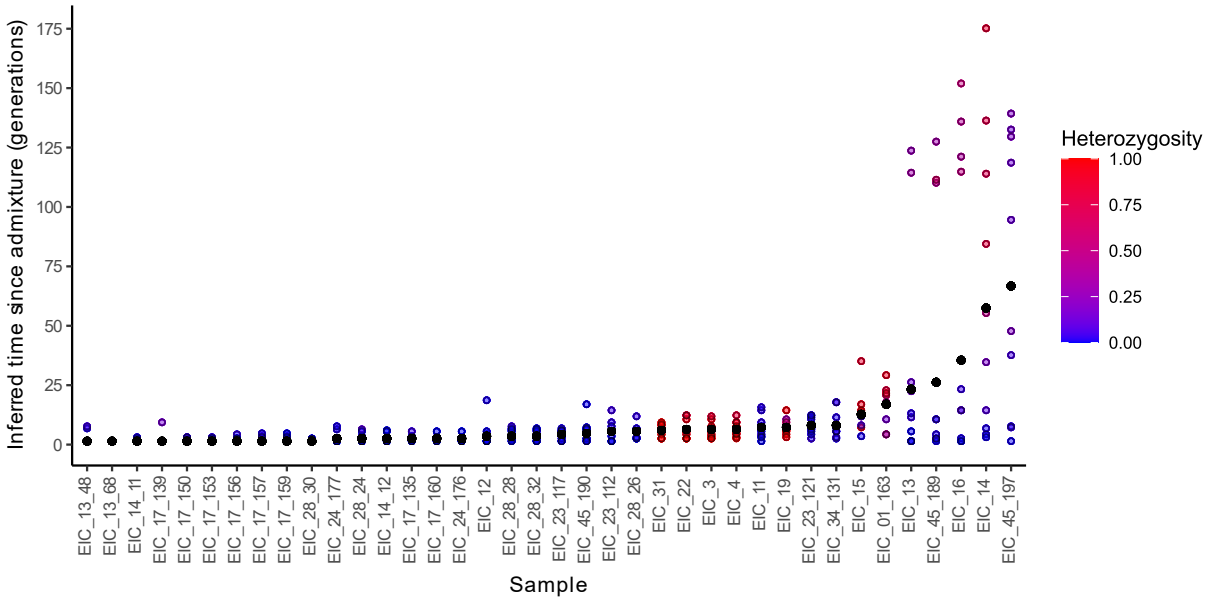


Figure S4.4. Inferred time since admixture for population EIC for all ten chromosomes per individual and corresponding heterozygosity. Samples are ordered by genome-wide time since admixture estimate, indicated by black dots.

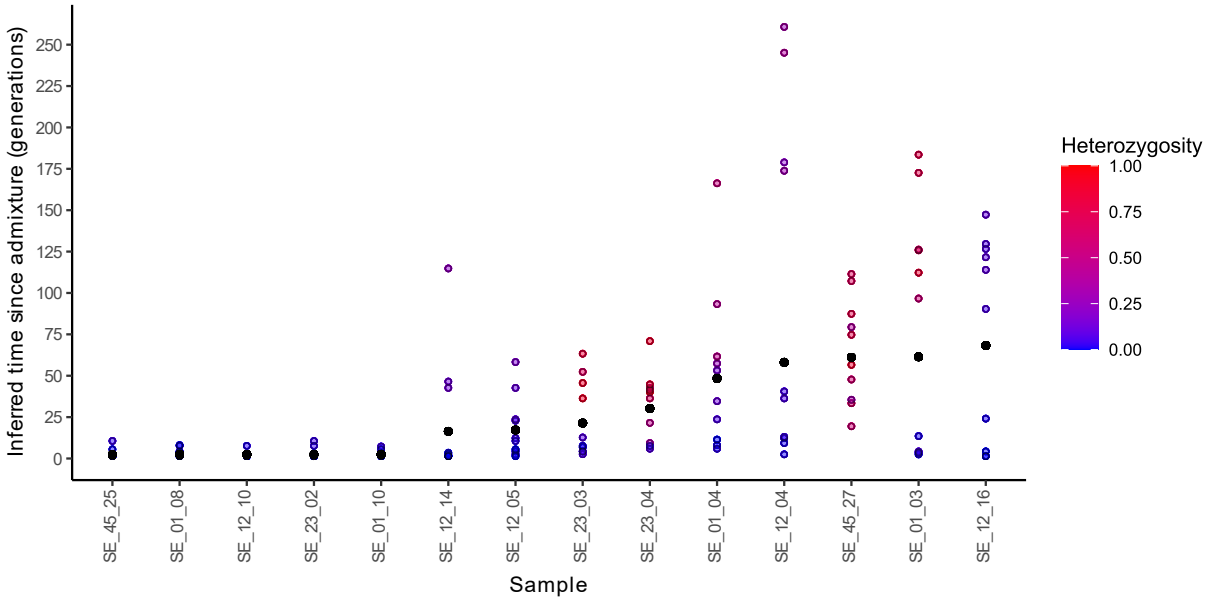


Figure S4.5. Inferred time since admixture for population SE for all ten chromosomes per individual and corresponding heterozygosity. Samples are ordered by genome-wide time since admixture estimate, indicated by black dots.

Figure S4.6. Local ancestry probability across the ten chromosomes for population EIC.

Figure S4.7. Local ancestry probability across the ten chromosomes for population SE.

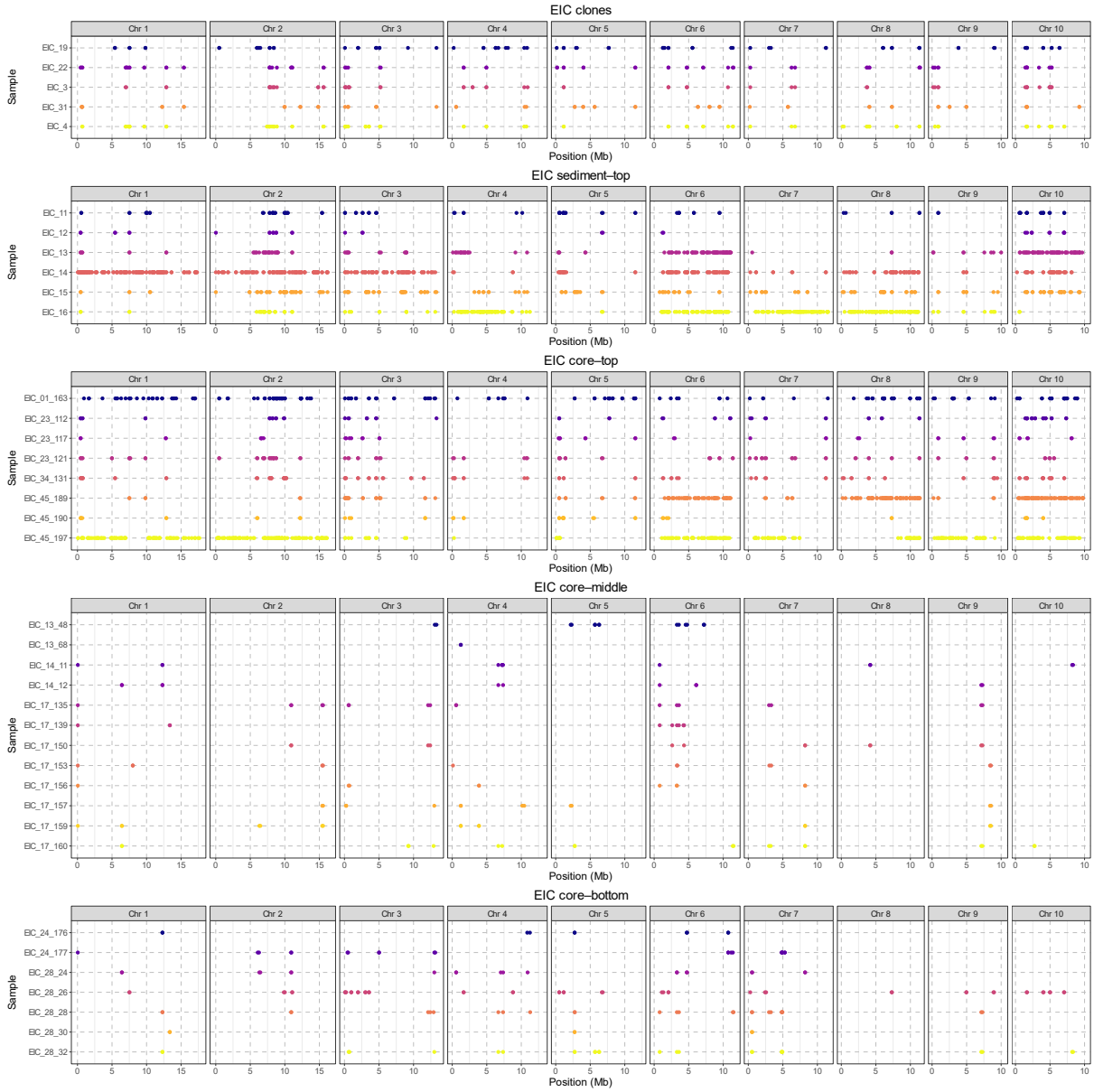


Figure S4.8. Junction position across the ten chromosomes separated into sample groups for population EIC.

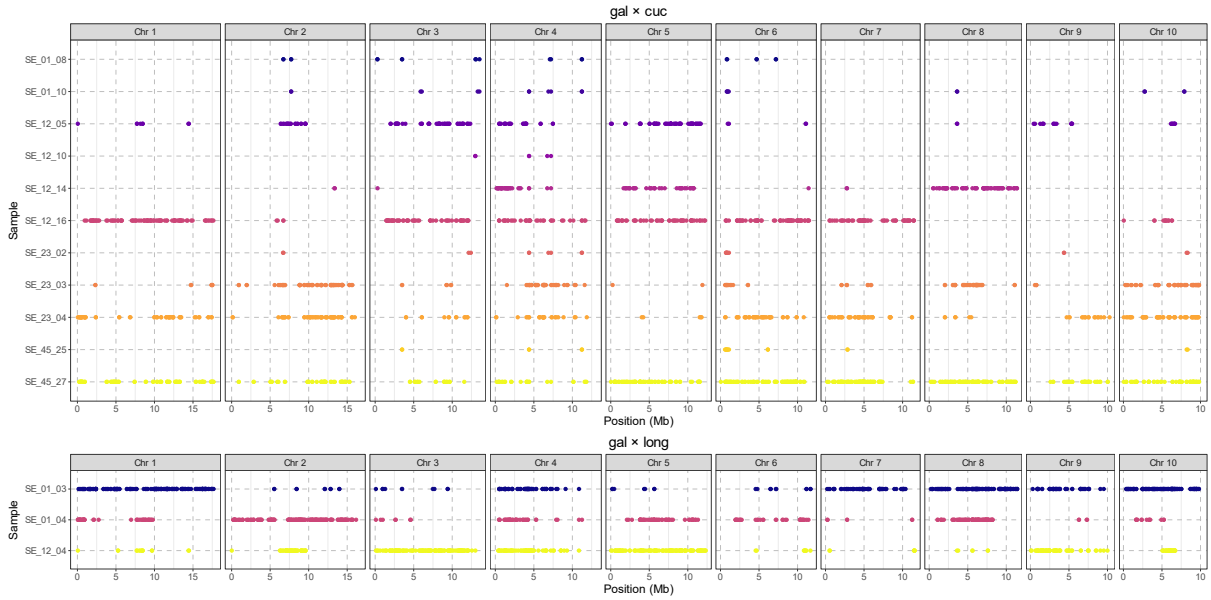


Figure S4.9. Junction position across the ten chromosomes separated into *D. galeata* × *D. cucullata and D. galeata* × *D. longispina* hybrids for population SE.

Supplementary tables

Table S4.1. Overview of the sample locations and basic characteristics of the lakes.

Location	Country	Latitude	Longitude	Max depth (m)	Surface area (km^2)	Trophic level	Stratification
Eichbaumsee	Germany	53.4852	10.1036	16	0.24	eutrophic	dimictic
Selenter See	Germany	54.3022	10.4400	36.8	21.25	mesotrophic	dimictic

Table S4.2. Overview of the 68 *Daphnia* samples included in this study and the sequencing protocol used. Sampled as: egg – genotype originates from ephippia isolated from sediment and genotyped directly; egg (hatched) – genotype originates from ephippia isolated from sediment cores and hatched in the lab; living (water column) – genotype was collected from the water column with a plankton net. Sediment layer: indicates the sediment depth range of the sediment core from which the egg was collected. DNA source: clone, unamplified – genotype was maintained in the lab and raised in bulk before DNA extraction; WGA – DNA was prepared by whole genome amplification of single resting eggs.

Sample name	Location	Sampled as	Sediment layer (cm)	DNA source	Species	Library preparation	ENA accession
EIC_19	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	NEB	ERS5080363, ERS14708726
EIC_22	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	NEB	ERS5080351, ERS14708727
EIC_3	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	Novogene & NEB	ERS5080352, ERS14708728
EIC_31	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	Novogene	ERS14708729
EIC_4	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	NEB	ERS5080353, ERS14708730
EIC_57	Eichbaumsee	living (water column)	/	Clone, unamplified	unknown	NEB	ERS5080354, ERS14708731
EIC_11	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080361
EIC_12	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080362
EIC_13	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080355
EIC_14	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080356
EIC_15	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080357
EIC_18	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080358
EIC_16	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080359
EIC_17	Eichbaumsee	egg	surface	WGA	unknown	NEB	ERS5080360
EIC_01_163	Eichbaumsee	egg	0–1	WGA	unknown	NEB	ERS14708732
EIC_23_112	Eichbaumsee	egg	2–3	WGA	unknown	NEB	ERS14708733
EIC_23_114	Eichbaumsee	egg	2–3	WGA	unknown	NEB	ERS14708734
EIC_23_117	Eichbaumsee	egg	2–3	WGA	unknown	NEB	ERS14708735
EIC_23_121	Eichbaumsee	egg	2–3	WGA	unknown	NEB	ERS14708736
EIC_34_131	Eichbaumsee	egg	3–4	WGA	unknown	NEB	ERS14708737
EIC_45_189	Eichbaumsee	egg	4–5	WGA	unknown	NEB	ERS14708738

EIC_45_190	Eichbaumsee	egg	4–5	WGA	unknown	NEB	ERS14708739
EIC_45_197	Eichbaumsee	egg	4–5	WGA	unknown	NEB	ERS14708740
EIC_13_48	Eichbaumsee	egg	13–14	WGA	unknown	NEB	ERS14708741
EIC_13_68	Eichbaumsee	egg	13–14	WGA	unknown	NEB	ERS14708742
EIC_14_11	Eichbaumsee	egg	14–15	WGA	unknown	NEB	ERS14708743
EIC_14_12	Eichbaumsee	egg	14–15	WGA	unknown	NEB	ERS14708744
EIC_17_135	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708745
EIC_17_139	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708746
EIC_17_150	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708747
EIC_17_153	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708748
EIC_17_156	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708749
EIC_17_157	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708750
EIC_17_159	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708751
EIC_17_160	Eichbaumsee	egg	17–18	WGA	unknown	NEB	ERS14708752
EIC_24_176	Eichbaumsee	egg	24–25	WGA	unknown	NEB	ERS14708753
EIC_24_177	Eichbaumsee	egg	24–25	WGA	unknown	NEB	ERS14708754
EIC_28_24	Eichbaumsee	egg	28–29	WGA	unknown	NEB	ERS14708755
EIC_28_26	Eichbaumsee	egg	28–29	WGA	unknown	NEB	ERS14708756
EIC_28_28	Eichbaumsee	egg	28–29	WGA	unknown	NEB	ERS14708757
EIC_28_30	Eichbaumsee	egg	28–29	WGA	unknown	NEB	ERS14708758
EIC_28_32	Eichbaumsee	egg	28–29	WGA	unknown	NEB	ERS14708759
SE_01_03	Selenter See	egg	0–1	WGA	unknown	Novogene	ERS5080332
SE_01_04	Selenter See	egg	0–1	WGA	unknown	Novogene	ERS5080333
SE_01_08	Selenter See	egg	0–1	WGA	unknown	Novogene	ERS14708760
SE_01_10	Selenter See	egg	0–1	WGA	unknown	Novogene	ERS14708761
SE_12_04	Selenter See	egg	1–2	WGA	unknown	Novogene	ERS5080331
SE_12_05	Selenter See	egg	1–2	WGA	unknown	Novogene	ERS5080337
SE_12_10	Selenter See	egg	1–2	WGA	unknown	Novogene	ERS5080338
SE_12_14	Selenter See	egg	1–2	WGA	unknown	Novogene	ERS14708762
SE_12_16	Selenter See	egg	1–2	WGA	unknown	Novogene	ERS14708763
SE_23_02	Selenter See	egg	2–3	WGA	unknown	Novogene	ERS5080334
SE_23_03	Selenter See	egg	2–3	WGA	unknown	Novogene	ERS5080335
SE_23_04	Selenter See	egg	2–3	WGA	unknown	Novogene	ERS5080336
SE_45_25	Selenter See	egg	4–5	WGA	unknown	Novogene	ERS14708764
SE_45_27	Selenter See	egg	4–5	WGA	unknown	Novogene	ERS14708765

ENDI44	Lago d'Endine	living (water column)	/	Clone, unamplified	gal	previous study	ERS5080368
G100	Tjeukemeer	living (water column)	/	Clone, unamplified	gal	previous study	ERS5080367
J2	Jordan Reservoir	egg (hatched)	/	Clone, unamplified	gal	previous study	ERS4993274
LC3_6	Lake Constance	egg (hatched)	/	Clone, unamplified	gal	previous study	ERS4993282
M5	Müggelsee	egg (hatched)	/	Clone, unamplified	gal	previous study	ERS5670685
BIEL6	Bielersee	living (water column)	/	Clone, unamplified	cuc	previous study	ERS5080370
SIHL4	Sihlsee	living (water column)	/	Clone, unamplified	cuc	previous study	ERS5080369
TS64	Thaler See	living (water column)	/	Clone, unamplified	cuc	previous study	ERS5080371
BODO_18_08	Lake Constance	living (water column)	/	Clone, unamplified	long	previous study	ERS5080374
GAR7	Lago di Garlate	living (water column)	/	Clone, unamplified	long	previous study	ERS5080373
KL11	Klostersee	living (water column)	/	Clone, unamplified	long	previous study & Novogene	ERS5080375, ERS14708766
NEUE02_08	Lac Neuchatel	living (water column)	/	Clone, unamplified	long	previous study	ERS5080372

Table S4.3. Contiguity metrics of the different assembly stages.

	# Sequences	Total length (bp)	N50
Published assembly			
Contigs	526	133,183,790	483,996
Scaffolds	346	133,304,630	756,671
Before manual curation			
Contigs	1,331	133,182,873	308,346
Scaffolds	489	133,638,130	68,195,500
After manual curation			
Contigs	1,700	133,181,685	266,000
Scaffolds	717	133,708,232	12,430,530

Table S4.4. Sequencing results for whole-genome data for the total raw, trimmed, and mapped reads and the read coverage.

Sample name	# Raw reads	# Trimmed reads	# Mapped reads	Mapped reads (%)	Coverage (×)
EIC_19	29,766,972	15,120,423	12,740,025	84.26%	12.14
EIC_22	36,089,700	19,249,150	17,004,385	88.34%	16.56
EIC_3	17,680,166	16,611,928	14,731,055	88.68%	15.60
EIC_31	48,443,612	29,767,990	21,114,902	70.93%	20.26
EIC_4	30,922,254	13,381,193	11,923,260	89.10%	11.27
EIC_57	31,835,176	16,692,404	13,229,154	79.25%	12.62
EIC_11	16,210,646	8,797,507	8,510,675	96.74%	7.96
EIC_12	19,121,834	2,796,407	2,594,281	92.77%	1.73
EIC_13	14,463,272	12,408,073	12,086,506	97.41%	12.48
EIC_14	16,917,644	12,054,388	11,760,401	97.56%	11.83
EIC_15	15,531,160	13,333,971	12,990,706	97.43%	13.42
EIC_18	14,952,714	10,070,564	9,367,638	93.02%	9.06
EIC_16	15,450,664	9,359,794	8,887,030	94.95%	8.51
EIC_17	15,392,966	12,549,996	12,080,619	96.26%	12.57
EIC_01_163	20,882,986	18,438,610	17,910,874	97.14%	18.99
EIC_23_112	20,308,844	18,442,430	17,427,688	94.50%	18.04
EIC_23_114	13,333,860	12,237,742	4,299,897	35.14%	3.83
EIC_23_117	15,531,648	13,429,570	12,749,299	94.93%	13.27
EIC_23_121	21,620,638	19,366,040	18,726,617	96.70%	19.54
EIC_34_131	19,955,938	17,587,074	16,645,051	94.64%	17.25
EIC_45_189	18,095,854	16,109,902	15,433,696	95.80%	16.13
EIC_45_190	15,882,552	14,089,622	11,141,418	79.08%	11.31
EIC_45_197	14,872,564	13,222,844	12,821,725	96.97%	13.53
EIC_13_48	17,381,378	14,997,298	14,571,468	97.16%	15.80
EIC_13_68	16,404,864	9,175,040	4,911,121	53.53%	4.36
EIC_14_11	16,786,576	14,370,358	14,053,940	97.80%	15.24
EIC_14_12	18,263,236	15,821,404	14,461,845	91.41%	15.70
EIC_17_135	17,287,616	14,412,000	14,115,211	97.94%	15.35
EIC_17_139	16,480,794	14,423,988	13,687,837	94.90%	14.85
EIC_17_150	25,775,904	23,309,066	22,210,923	95.29%	24.20
EIC_17_153	15,568,752	13,745,018	12,590,744	91.60%	13.50
EIC_17_156	24,081,004	20,554,278	20,015,826	97.38%	21.71
EIC_17_157	24,453,928	22,227,152	21,601,330	97.18%	23.49

EIC_17_159	23,199,950	21,250,506	20,697,370	97.40%	22.51
EIC_17_160	8,722,660	7,681,072	7,506,756	97.73%	8.12
EIC_24_176	105,788,320	93,389,112	88,603,427	94.88%	95.37
EIC_24_177	14,783,840	12,942,944	12,644,961	97.70%	13.69
EIC_28_24	16,848,866	14,605,350	10,513,571	71.98%	10.93
EIC_28_26	18,037,154	14,643,700	14,003,986	95.63%	14.49
EIC_28_28	18,263,424	15,536,874	14,606,545	94.01%	15.75
EIC_28_30	18,463,972	14,152,900	11,986,386	84.69%	12.54
EIC_28_32	22,766,444	9,049,114	8,794,769	97.19%	9.33
SE_01_03	27,737,288	22,465,594	21,552,959	95.94%	22.71
SE_01_04	23,550,114	21,910,332	21,298,998	97.21%	23.02
SE_01_08	17,602,464	16,727,400	16,245,072	97.12%	17.78
SE_01_10	18,327,884	17,396,112	16,926,810	97.30%	18.48
SE_12_04	42,684,512	41,338,814	13,098,321	31.69%	13.00
SE_12_05	20,708,746	19,447,996	19,180,655	98.63%	21.06
SE_12_10	22,383,236	21,046,994	20,457,532	97.20%	22.43
SE_12_14	19,080,266	17,911,846	17,140,930	95.70%	18.55
SE_12_16	16,871,550	16,080,800	15,654,187	97.35%	17.13
SE_23_02	19,136,852	18,015,214	17,534,200	97.33%	19.18
SE_23_03	19,537,960	18,398,560	17,893,821	97.26%	19.41
SE_23_04	27,909,232	26,384,906	24,918,200	94.44%	26.96
SE_45_25	18,689,358	17,696,784	17,016,996	96.16%	18.24
SE_45_27	21,330,852	20,271,950	19,630,813	96.84%	21.23
J2	79,048,948	73,020,526	65,159,511	89.23%	69.93
LC3_6	90,918,942	85,065,214	61,806,255	72.66%	66.03
M5	77,594,732	64,988,122	54,985,625	84.61%	56.87
G100	39,015,978	31,495,250	29,785,307	94.57%	29.77
ENDI44	41,567,550	39,657,286	37,973,543	95.75%	41.02
SIHL4	55,802,822	43,669,966	40,869,011	93.59%	38.82
BIEL6	44,908,864	42,806,350	39,962,490	93.36%	41.34
TS64	95,251,354	77,115,502	71,784,170	93.09%	68.77
NEUE02_08	91,181,974	76,574,168	72,067,493	94.11%	68.53
GAR7	66,935,478	52,166,782	48,726,762	93.41%	45.79
BODO_18_08	64,818,020	62,076,398	58,615,703	94.43%	60.72
KL11	52,141,742	33,510,748	24,224,774	72.29%	23.18

Table S4.5. Inferred time since admixture for all hard ancestry call thresholds and the jointly estimated time since admixture using Ancestry HMM.

		Threshold					
Sample name	**Group**	**$1-10^{-6}$**	**$1-10^{-5}$**	**$1-10^{-4}$**	**$1-10^{-3}$**	**$1-10^{-2}$**	**$1-10^{-1}$**
EIC_19	Clones	3.5	4.3	6.4	7.3	9.4	14.4
EIC_22	Clones	4.3	4.3	4.3	6.4	9.4	13.6
EIC_3	Clones	3.5	3.5	4.3	6.4	6.8	9.4
EIC_31	Clones	3.1	3.5	3.5	6.0	6.8	8.1
EIC_4	Clones	4.3	4.3	4.3	6.4	9.4	15.7
EIC_11	Sediment–top	2.5	3.1	4.7	7.3	11.9	18.2
EIC_12	Sediment–top	2.6	2.6	2.6	3.5	6.8	14.4
EIC_13	Sediment–top	16.5	17.0	19.5	23.3	29.2	40.6

EIC_14	Sediment–top	27.9	35.1	47.8	57.5	75.6	96.7
EIC_15	Sediment–top	3.5	4.3	7.7	12.7	21.6	34.7
EIC_16	Sediment–top	17.8	24.6	31.3	35.5	44.0	59.6
EIC_01_163	Core–top	10.6	11.9	14.4	17.0	20.8	29.2
EIC_23_112	Core–top	2.6	3.1	4.3	5.6	8.5	17.4
EIC_23_117	Core–top	2.6	2.6	2.6	4.3	6.8	13.2
EIC_23_121	Core–top	3.5	4.3	6.4	8.1	12.3	17.4
EIC_34_131	Core–top	2.6	3.1	4.3	8.1	12.3	18.6
EIC_45_189	Core–top	16.5	18.2	22.4	26.2	31.3	43.5
EIC_45_190	Core–top	2.6	2.6	3.1	4.7	7.7	13.6
EIC_45_197	Core–top	54.9	57.5	63.4	66.7	70.5	84.5
EIC_13_48	Core–middle	1.4	1.4	1.4	1.4	2.5	3.5
EIC_13_68	Core–middle	1.4	1.4	1.4	1.4	1.4	1.4
EIC_14_11	Core–middle	1.4	1.4	1.4	1.4	2.6	2.6
EIC_14_12	Core–middle	2.6	2.6	2.6	2.6	2.6	2.6
EIC_17_135	Core–middle	1.4	2.6	2.6	2.6	2.6	3.1
EIC_17_139	Core–middle	1.4	1.4	1.4	1.4	1.4	2.6
EIC_17_150	Core–middle	1.4	1.4	1.4	1.4	1.4	2.6
EIC_17_153	Core–middle	1.4	1.4	1.4	1.4	1.4	2.6
EIC_17_156	Core–middle	1.4	1.4	1.4	1.4	1.4	2.5
EIC_17_157	Core–middle	1.4	1.4	1.4	1.4	1.4	2.6
EIC_17_159	Core–middle	1.4	1.4	1.4	1.4	2.5	2.6
EIC_17_160	Core–middle	2.6	2.6	2.6	2.6	2.6	3.5
EIC_24_176	Core–bottom	1.4	1.4	1.4	2.6	2.6	3.5
EIC_24_177	Core–bottom	1.4	1.4	2.5	2.5	2.6	2.6
EIC_28_24	Core–bottom	1.4	1.4	2.5	2.5	2.6	3.1
EIC_28_26	Core–bottom	2.6	3.5	4.3	5.6	6.4	9.4
EIC_28_28	Core–bottom	3.5	3.5	3.5	3.5	4.3	5.6
EIC_28_30	Core–bottom	1.4	1.4	1.4	1.4	2.6	2.6
EIC_28_32	Core–bottom	2.6	3.1	3.1	3.5	4.3	4.3
SE_01_03	SE core	37.6	44.8	52.4	61.7	80.2	109.8
SE_01_04	SE core	25.4	31.3	37.6	48.6	57.5	77.3
SE_12_04	SE core	25.4	38.5	46.5	58.3	73.5	104.3
SE_01_08	SE core	1.4	1.4	1.4	2.6	3.1	3.5
SE_01_10	SE core	1.4	1.4	1.4	2.6	2.6	3.5
SE_12_05	SE core	8.5	10.6	14.4	17.4	21.6	28.8
SE_12_10	SE core	1.4	1.4	2.6	2.6	2.6	2.6
SE_12_14	SE core	11.9	12.3	14.4	16.5	18.2	22.9
SE_12_16	SE core	64.6	65.9	65.9	68.4	69.7	73.5
SE_23_02	SE core	1.4	1.4	2.6	2.6	2.6	3.5
SE_23_03	SE core	15.7	16.5	19.5	21.6	25.4	33.4
SE_23_04	SE core	22.9	24.1	27.5	30.5	35.5	44.0
SE_45_25	SE core	1.4	1.4	1.4	2.5	2.6	2.6
SE_45_27	SE core	44.0	48.6	51.6	61.3	68.4	87.4

Acknowledgments

I'm incredibly grateful to all the people who have supported and motivated me on this journey and made this thesis possible:

First of all, I want to thank my supervisor, **Mathilde Cordellier**. Thank you for believing in me throughout my entire scientific journey and for giving me the confidence to step out of my comfort zone. I've learned so much from you and I'm so grateful for your dedication, encouragement, and guidance at every step of the way, even under sometimes pretty difficult circumstances.

I want to thank my unofficial second supervisor **Markus Möst** for his time, brilliant advice, and close attention to detail, and especially for stepping in at exactly the right time to support me in finishing my thesis.

Many thanks to **Susanne Dobler** for always taking an interest in my work and allowing me to stay for the time I needed.

I'm grateful to **Tilman Schell** for patiently teaching me almost everything there is to know about genome assembly.

Many thanks to **Thijs Janzen** for taking our genomic data and rough ideas and turning them into something amazing.

I would also like to thank all the people who provided us with *Daphnia* and sediment samples and helped us on more or less adventurous boat trips. In addition, I'm also grateful for the many hands that helped me sort through thousands of resting eggs.

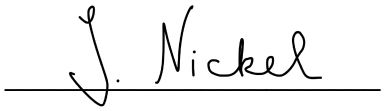
Thanks to the former and current members of the Population Genomics and Molecular Ecology groups throughout my Bachelor, Master, and then PhD years for their company, helpful advice, and helping me see things from a fresh perspective. Special thanks to my office mate **Jenni** for her support and excellent company.

Most of all, I want to thank my mother **Claudia** and my little brother **Ole**. Thank you for your unwavering love and for supporting me in so many ways. I couldn't have done it without you.

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 20.09.2023



Jana Nickel