Comparative and population transcriptomics of *Daphnia galeata*

Dissertation submitted with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Biology
Universität Hamburg

Submitted by
Suda Parimala Ravindran
July 2018

Thesis Evaluators:

1. Prof. Dr. Mathilde Cordellier
2. Dr. Luisa Orsini

"Learning gives creativity
Creativity leads to thinking
Thinking provides knowledge
Knowledge makes you great."

— A.P.J. Abdul Kalam

To

Daddy and Amma

... for their innumerable sacrifices, unconditional love, encouragement and trust since my existence

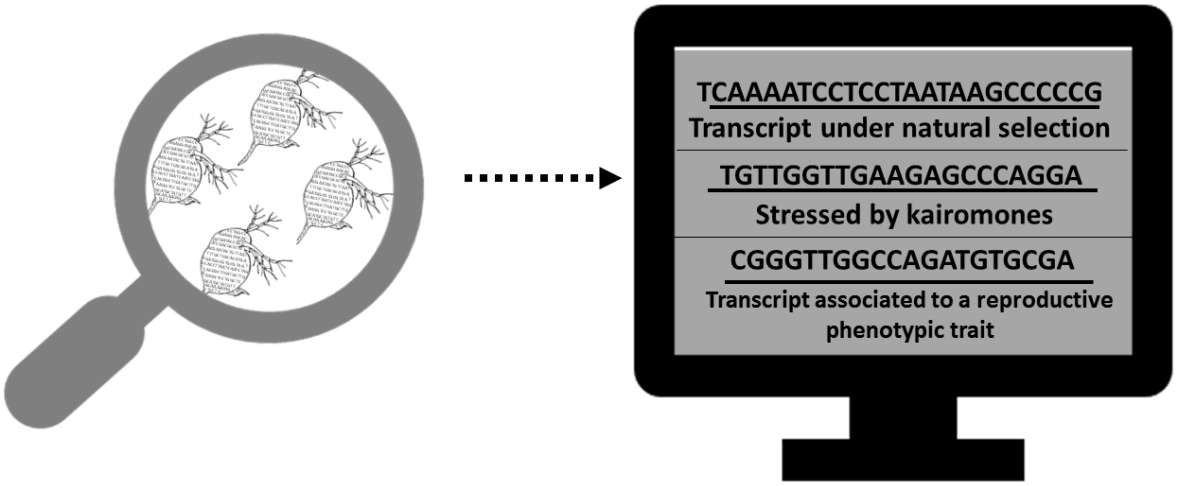

TCAAAATCCTCCTAATAAGCCCCCG
Transcript under natural selection
TGTTGGTTGAAGAGCCCAGGA
Stressed by kairomones
CGGGTTGGCCAGATGTGCGA
Transcript associated to a reproductive phenotypic trait

Table of Contents

List of Figures

List of Tables

List of abbreviations

AFR	Age at First Reproduction
ANOVA	Analysis of Variance
BLAST	Basic Local Alignment Search Tool
BP	Biological Process
CC	Cellular Component
db	Database
DET	Differentially Expressed Transcript(s)
DGC	*Daphnia* Genomics Consortium
FC	Fold Change
GATK	Genome Analysis ToolKit
GO	Gene Ontology
GS	Gene Significances
GWAS	Genome Wide Association Studies
HTML	Hyper Text Markup Language
MAFFT	Multiple Alignment using Fast Fourier Transform
ME	Module Eigengene
MF	Molecular Function
PCA	Principal Component Analysis
PHP	Hypertext Preprocessor
SGR	Somatic Growth Rate
SNP	Single Nucleotide Polymorphism(s)
SQL	Structured Query Language
TOM	Topological Overlap Matrix
VCF	Variant Call Format
WGCNA	Weighted Gene Co-expression Network Analysis
NCBI	National Center for Biotechnology Information
EMBOSS	European Molecular Biology Open Software Suite
SAM	Sequence Alignment/Map format

Abstract

Natural environments are increasingly fragmented through habitat destruction and organisms shift their geographical distribution by altering their growth and reproduction patterns to persist in the habitat. Organisms are subjected to severe selection pressures and they undergo rapid evolutionary changes to avoid extinction. This is particularly evident in aquatic environments especially for species like *Daphnia*. Therefore, understanding the genetic basis of local adaptation and phenotypic variation and their interplay with the environment are important factors to predict the consequences of ongoing climate change. *Daphnia* are cyclical parthenogens and they produce resting eggs which make them a well-suited model organism for this purpose.

The aim of the thesis was to assess the impact of transcriptomic basis of local adaptation and phenotypic variation in four European *D. galeata* populations and identify their ecological roles. To this aim, I used the *D. galeata* transcriptome data to identify transcripts potentially under local adaptation at the regulatory and sequence level. I revealed candidate transcripts under local adaptation which is influenced by the interplay of forces such as drift, selection and gene flow. Furthermore, I performed a functional annotation of *D. galeata* transcripts. However, several transcripts identified in *D. galeata* were “hypothetical” in function. To be able to assess the ecological role and improve the existing annotations of these transcripts, I employed a data mining approach and developed a database based on the expression profiles in *Daphnia* species from several previous studies. This meta-analysis helped to improve the existing functional annotation of *Daphnia* by linking them to their ecological roles. Finally, to reveal the transcriptomic basis of phenotypic variation, I performed an association analysis with sequence polymorphisms and ten life-history traits data in the absence and presence of predator-kairomones obtained from another experiment. This study revealed a transcriptomic basis for two life-history traits and the role of plasticity in determining adaptations was inferred.

In conclusion, I detected signs of local adaptation at the transcriptomic level and the association analysis allowed assessing the interplay between the genotype, phenotype and environment. I also improved the existing functional annotation of *Daphnia*

genes/transcripts by developing the *Daphnia* stressor database. However, further information and work are required to predict the consequences of *Daphnia* species accurately in the light of environmental fluctuations.

I. General introduction

"*As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be naturally selected. From the strong principle of inheritance, any selected variety will tend to propagate its new and modified form.*"

-Charles Darwin, *On the origin of species*, 1859

Natural habitats are subjected to environmental constraints due to habitat fragmentation which affects all levels of biodiversity: individuals, populations, species, ecological networks and ecosystems (reviewed in Bellard *et al.* 2012). Individuals in a population are affected by mutations and recombination and/or they exhibit plastic responses to persist in the long run. At the population level, habitat fragmentation decreases genetic diversity in populations due to rapid migration and selection (Botkin *et al.* 2007a). Beyond population level, the various effects of environmental constraints are seen at the community level (Gilman *et al.* 2010; Walther 2010). Community structure is modified by factors of interspecific relationships such as competition, predation, parasitism or mutualism (Lafferty 2009; Walther 2010; Yang & Rudolf 2010). Thus, habitat fragmentation has potentially strong implications on natural populations and their biodiversity, leading to the most extreme outcomes such as species extinction (Rafferty & Ives 2011).

The effects of habitat fragmentation are particularly evident in freshwater ecosystems as they are constantly under overexploitation of water and organisms, pollution and invasion of exotic species, all of which are due to human activities. It is important to understand how aquatic individuals, populations or species avoid these effects and persist in the long run by responding adaptively. Therefore, understanding the molecular basis of evolutionary adaptation in response to environmental changes is not only one of the major goals in biology but will also aid in the prediction of environmental changes for populations and species from the past and future more precisely (Pardo-Diaz *et al.* 2015).

I-1 Genetic versus plastic adaptive responses

Populations adapt to their environments through natural selection (Darwin 1859). Selection pressures differs spatially across the species distribution range when populations inhabit heterogeneous environments (Hedrick 2006). Thus, as a response (Figure 1), populations (i) become locally adapted, where species can adapt genetically to their environments through selection of new genotypes or mutations (Salamin *et al.* 2010) or (ii) evolve phenotypic plasticity, which provides a short-term adaptive response within an individual's lifetime (Charmantier *et al.* 2008), to persist in the environment.

I-1.1 Genetic variation

Genetic diversity is important for populations to adapt to changing environments. Genetic variation can occur at the (a) individual level and/or (b) population-level. Individuals in a population undergo genetic mutations that can provide either a survival advantage or sometimes become maladaptive. Other possible explanations for genetic variation at the individual level are processes such as meiosis (independent assortment and crossing over) and gamete combinations.

Population structure, which is the pattern of genetic variation among populations is influenced by three main processes: (a) gene flow and dispersal; (b) random genetic drift and (c) natural selection.

I-1.1a Gene flow and dispersal

Genetic flow is the transfer of alleles between different populations through the process of migration, leading to changes in allele frequencies of the populations. In a heterogeneous environment, gene flow is one of the major factors affecting local adaptation. Gene flow introduces new alleles that might be beneficial and can displace the old alleles, if they are better adapted in the new habitat. In locally adapted populations, high rates of gene flow restrict adaptive divergence (Lind *et al.* 2011) and may also result in the evolution of phenotypic plasticity (DeWitt & Scheiner 2004). Low rates of gene flow are often associated to population extinction as seen in populations of the butterfly *Melitaea cinxia*, where low survival rates and fecundity of the individuals is observed

(Saccheri *et al.* 1998). Thus, gene flow restricts isolation between populations thereby preventing them from diverging due to drift or selection.

Dispersal is another mechanism by which individuals or propagules that show potential for gene flow across space migrate to new environments. Organisms disperse either because of their innate tendency to disperse or maybe associated to other processes such as foraging and mate searching (Bonte & Dahirel 2017; Burgess *et al.* 2016; Van Dyck & Baguette 2005). Dispersal can be active or passive. In actively dispersing organisms, an individual's capacity to disperse depends on its morphology, size or endurance (Saastamoinen *et al.* 2018). Dispersal can often be seen in plants, where seeds disperse through wind and their dispersal ability depends on various factors like overall size of the plant, height of seed release, seed size and its morphology (Soons *et al.* 2004; Thomson *et al.* 2010). Passive dispersal is often vectored by agents such as wind, water, animals, gravity or human interference (Saastamoinen *et al.* 2018). Passive dispersal is also found in animal taxa (e.g., Houck & OConnor 1991) and in aquatic species as seen for example in *Daphnia* (Pietrzak 2006; van de Meutter *et al.* 2008). Both active and passive dispersal are correlated to phenotypic characteristics such as morphology or life-history traits such as reproduction (Bonte & Dahirel 2017; Stevens *et al.* 2012; Stevens *et al.* 2014), resulting in co-evolution when the correlations have a genetic basis. Further, environmental conditions also can strongly influence dispersal capabilities (Bowler & Benton 2005; Matthysen 2005), and wind forces have an impact on the distances of seed dispersal (Greene 2005).

I-1.1b Genetic drift

Genetic drift, which is the random change in allele frequencies over generations, is another important mechanism of evolution. Although genetic drift can occur in populations of all sizes, its effect is the strongest in small populations. Genetic drift occurs when population size decreases due to changes in environment (bottleneck effect) or when individuals in a population emigrate to new locations and establish their colony (founder effect).

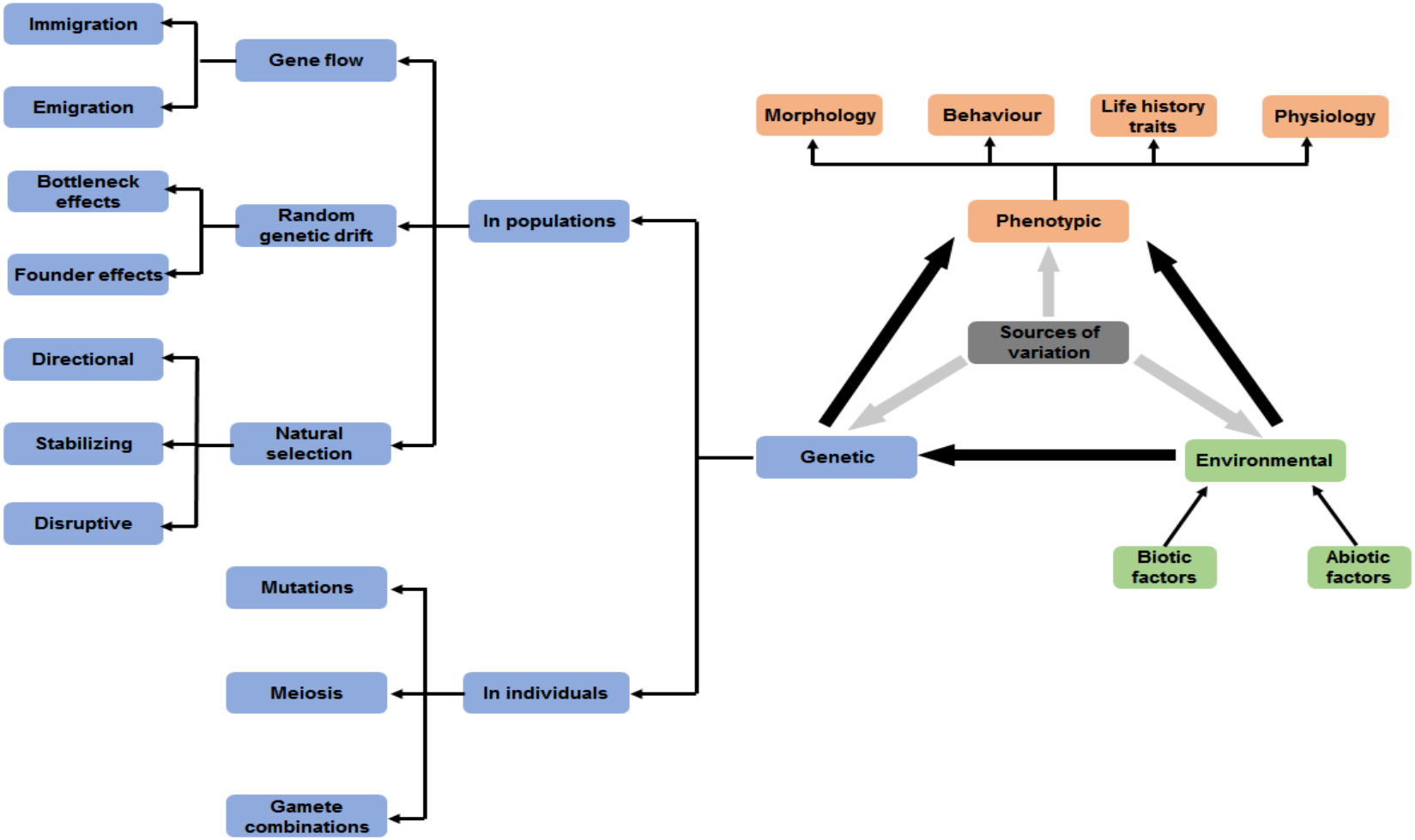


Figure 1: Consequences of variation at different levels.

Shown here are the three different sources of variation: environment (in green), phenotype (in orange) and genotype (in blue). Grey arrows indicate the three different sources of variation viz., environmental, genetic and phenotypic variation. Black arrows represent the interaction between the three sources of variation.

A ***bottleneck effect*** is an extreme example of genetic drift and is the non-selective change in allele frequencies that happen by random chance, thereby reducing the genetic diversity within populations. A classic example of a bottleneck event is the northern elephant seal, which was predominantly hunted in the 18th century for their oil-producing blubbers. However, a small colony of northern elephant seals (about 20 to 100 individuals) survived on the Guadalupe Island and this dramatic shrinking of the populations resulted in low genetic diversity of the individuals, and this phenomenon is called a genetic bottleneck (Hoelzel *et al.* 2002). Several studies have shown the correlation between genetic variation and fitness measures (e.g. Koehn *et al.* 1988; Leary *et al.* 1985; Soulé 1979) due to the impact of inbreeding depression. A study on song sparrows populations that have undergone bottleneck events validated that the inbreeding coefficient was significantly lower for birds that survived a population crash than those birds that did not survive (Keller *et al.* 1994). Thus, the loss of genetic diversity during population bottlenecks reduces the evolutionary potential of a population in the light of changing environments (Fisher 1930).

There are several methods to measure genetic variation and its loss and each of them have their advantages and pitfalls. First is the effect of random sampling on subsequent gene frequencies. Alleles with higher frequencies tend to have a higher possibility of survival; however, the presence of a rare allele can result in being the most common allele in the post-bottleneck population (Hoelzel 1999). But in general, rare alleles have fewer chances of survival and thus a reduction in polymorphism levels and allele frequency redistribution is observed. This bias in allele frequencies can be detected and modelled as a 'bottleneck signature' (Luikart *et al.* 1998a).

A second explanation is based on the heterozygosity in diploid organisms. Heterozygosity is a widely used measure and it confounds the effect of population size on genetic variation (Allendorf 1986). Heterozygosity measures are one of the best estimates of genetic variation in populations that are in Hardy-Weinberg equilibrium (Nei & Roychoudhury 1974). Low levels of heterozygosity are caused by bottleneck events in the habitat. For example, a study on moths by Menken (1987) revealed low average heterozygosity levels and the authors attribute bottleneck events in the habitat as one of the factors for such observation. However, heterozygosity measures have the

disadvantage of being insensitive to the actual number of genotypes in a population. Therefore, using only heterozygosity measures to predict the effects of a small population size on genetic variation may result in a biased outlook.

The third factor that explains loss of genetic variation is the disruption of pleiotropic interactions. High levels of morphometric variation are observed in inbred populations (Waddington 1957), and an inverse relationship between morphological variation and allozyme heterozygosity has been observed in several species (Fleischer *et al.* 1991; Leary *et al.* 1985; Mitton 1978). Population bottleneck events can reduce quantitative trait variation due to a reduction in additive genetic variance (Carson & Templeton 1984; Templeton 1979). On the contrary, bottleneck events can increase the variation in quantitative traits when the expressed variance is due to non-additive interaction such as epistasis or dominance (Bryant *et al.* 1986; Goodnight 1988). This pattern is observed for example in an experiment with houseflies (Bryant *et al.* 1986), where four out of eight morphological characters showed increased variation due to bottleneck events. Thus, bottleneck events have a major effect on allelic diversity and therefore might have a significant effect on long-term adaptation and survival of the populations or species.

Founder effects describe the loss of allelic variation that occurs when a very small number of individuals in a population shift to new locations and establish their colony. This effect might cause the newly founded population to differ considerably from the source population. Founder effects have been a theoretically well-studied mechanism in vertebrate speciation (Boake & Gavrilets 1998; Gavrilets & Hastings 1996; Rundle *et al.* 2017; Turelli *et al.* 2001). However, few studies have tested this phenomenon empirically and have shown that founder effects are rare (Walsh *et al.* 2005). For example, two studies (Johnson *et al.* 1996; Meyer *et al.* 1990) on cichlid fishes that originated in the African Rift Lakes in <1 million years ago, showed that extensive intraspecific variation in major histocompatibility complex (MHC) suggested that their speciation was not due to founder effects (Klein *et al.* 1993). Another study on three-spined sticklebacks (McKinnon & Rundle 2002) found no evidence for speciation due to founder effects, despite showing evidence for rapid radiation. On the contrary, founder effects have shown to be the primary mode of speciation among Darwin's finches (Vincek *et al.* 1997). Rapid colonization and fast population growth by a small number of founders can result in strong founder events as

seen for example in Ethiopian *Daphnia* (Haileselasie *et al.* 2018). Thus, founder effects increase the allele frequency of populations and can reach even higher frequencies due to strong genetic drift, which occurs while the population size is still small (Slatkin 2004).

To sum up, both bottleneck and founder events have profound effects on the evolutionary adaptation of a population. Genetic drift is arguably the most powerful evolutionary force in small populations and can change the allele frequencies of populations by random chance. However, allele frequencies of populations can also change by differential reproductive success and this mechanism is known as natural selection.

I-1.1c Natural selection

Natural selection is another important mechanism of evolution that occurs when individuals with specific genotypes are most likely to survive and reproduce than individuals with another genetic makeup, thereby acquiring adaptations and pass down the alleles to their offspring. Only populations with favourable adaptive features will survive and evolve through speciation. Three types of natural selection exist: (a) stabilizing; (b) directional and (c) diversifying selection, which all can either increase, decrease or shift the genetic differentiation in populations.

Directional selection is a mode of natural selection where a single genotype is favoured thereby leading to shifts in allele frequency in one direction. There are abundant evidence for directional selection on life-history traits and morphology in several organisms (Hereford *et al.* 2007; Kingsolver & Diamond 2011; Kingsolver *et al.* 2001). For example, a study on East African cichlid fishes revealed that cichlid jaws and teeth have evolved due to strong directional selection on chromosomal regions that encode functionally relevant traits of craniofacial skeleton (Albertson *et al.* 2003). Studies claim that to produce rapid micro-evolutionary adaptations, directional selection alone is sufficient (Grant & Grant 1989; Hendry & Kinnison 1999; Reznick & Ghalambor 2001). There are several reasons that can explain why directional selection is commonly occurring. First, the lack of genetic diversity prevents evolutionary responses to selection. Several quantitative genetic studies reveal that there is an abundant heritable variation for phenotypic traits (e.g. Mousseau & Roff 1987), and lower heritability for some

phenotypic traits and study species (e.g. Kellermann *et al.* 2009). Additionally, multivariate genetic constraints may lead to reduced adaptive responses to selection, even in the presence of high genetic variation of individual traits (Walsh & Blows 2009). Secondly, trade-offs may be common among different components of fitness-related traits, for example, a trait value increasing the rates of survival might also decrease mating success or fecundity, thereby increasing the scope for directional selection (Kingsolver & Diamond 2011). Thirdly, genetic and phenotypic correlations among traits might cause indirect and correlated selection thereby counterbalancing directional and indirect selection on a correlated trait (Lande & Arnold 1983). Lastly, directional selection on a trait may alternate in time or space, thereby reducing the effects of selection (Siepielski *et al.* 2009).

A ***stabilizing selection*** is another mode of natural selection where an average phenotype is favoured and stabilized than the extreme phenotypes and might lead to a decrease in genetic variation. Stabilizing selection has much stronger effects on natural populations compared to what is predicted in theoretical models (Johnson & Barton 2005). Further, stabilizing selection is the cause of stasis over long evolutionary timescales (Estes & Arnold 2007). This pervasiveness of genetic diversity despite stabilizing selection has no valid explanations (Johnson & Barton 2005). Although stabilizing selection is assumed to be common, there are very little experimental evidence that support this theory (Geber & Griffen 2003; Kingsolver *et al.* 2001; Smarda *et al.* 2010; Travis 1989).

Disruptive selection, also known as ***diversifying selection***, is a mode of natural selection where extreme trait values are favoured, each having specific advantages and leading to an increase in genetic variance in the populations. The intermediate phenotypes become less fit than the extreme phenotypes and it is unlikely that they would persist in the habitat. Several theoretical studies support the importance of disruptive selection in sympatric speciation (Abrams *et al.* 2008; Bolnick & Fitzpatrick 2007; Dieckmann & Doebeli 1999; Doebeli *et al.* 2007; Rueffler *et al.* 2006) and empirical evidence supporting this hypothesis has been observed in multiple contexts such as: (i) To increase the divergence potential, disruptive selection can act between the competing species as seen in *Spea* tadpoles (Pfennig *et al.* 2007) and Darwin's finches (Grant & Grant 2006); (ii) Disruptive selection can result from competition between hybridized individuals and their parental forms as seen in threespine stickleback (Schluter 1994, 2003); (iii)

disruptive selection could occur in randomly mating bimodal populations as seen in African *Pyrenestes* finches (Smith 1993). Therefore, disruptive selection is a common phenomenon occurring in nature and they play an important role in adaptive radiation.

Thus, all the above-mentioned forces such as genetic drift, natural selection, dispersal and gene flow can influence how populations adapt to survive in the long run. One of the fundamental questions in biology is to understand what makes individuals, populations and species different from each other. Genetic variation and phenotypic variation are two fundamentally different levels of variation and understanding their interplay with the environment is essential to predict evolutionary adaptation. Therefore, it is important to understand the genetic basis of phenotypic variation.

I-1.2 Phenotypic variation

Phenotypic plasticity is defined as "the ability of individual genotypes to produce different phenotype when exposed to different environmental conditions" (Pigliucci *et al.* 2006). This mechanism involves both inter-and intra-specific variation in morphological, behavioural, physiological and life-history traits (Botkin *et al.* 2007b; Chevin *et al.* 2010). Some of the well-known and remarkable examples of plasticity are: environmental sex-determination in reptiles (e.g. Janzen & Phillips 2006) and predator-induced plastic polyphenism in cladocerans (e.g. Laforsch & Tollrian 2004b), and changes like acclimatization are also part of an organism's plastic responses (Fusco & Minelli 2010). Phenotypic plasticity is considered as a trait by itself (de Jong 2005), and is usually viewed as the interaction between the genotype and the environment (Fusco & Minelli 2010).

Phenotypic plasticity alters genetic diversity available to selection (Gibson & Dworkin 2004; Hermisson & Wagner 2004; Le Rouzic & Carlborg 2008; McGuigan & Sgro 2009; Paaby & Rockman 2014). There are two ways in which plasticity contributes to genetic variation. Firstly, homeostatic mechanisms which are expended, buffers the effects of novel mutations leading to reduced genetic constraints (Moczek 2008). For example, the heat shock protein Hsp90 canalizes the protein folding outcomes during regulatory mechanism and stabilizes proteins in the cells, thereby reducing the effects of genetic and environmental changes and endorsing the accumulation of genetic variation (Queitsch *et al.* 2002). Secondly, phenotype plasticity may lead to conditional expression

of genetic diversity in a few individuals or populations (Des Marais *et al.* 2013; Snell-Rood *et al.* 2011), which can increase the polymorphisms at specific loci (Kawecki 1994; Lahti *et al.* 2009; Snell-Rood *et al.* 2010; Van Dyken & Wade 2010). Therefore, evaluating these factors not only helps to understand the evolutionary basis of phenotypic variability but also adaptation to novel environments (Ghalambor *et al.* 2007) and the divergence of speciation and adaptive radiations (Fitzpatrick 2012; Pfennig *et al.* 2010; Thibert-Plante & Hendry 2011).

I-1.2a Inducible defences

Inducible defences, which are phenotypic changes triggered by chemical cues from biotic agents, such as predators, parasites or pathogens, are a special type of phenotypic plasticity. Inducible defences are a widespread mechanism and can be expressed at the morphological, behavioural and life-history levels (e.g. Harvell 1990; Kishida *et al.* 2009). They are found in bacteria, protozoa, insects, molluscs, amphibians, crustaceans and mammals.

Predator-induced life-history changes comprise trade-offs between reproduction and somatic growth (Weiss *et al.* 2012), thereby increasing the chances of reproduction and decreased predatory responses (Lass & Spaak 2003). For example, inducible defences allow the prey to overcome high predation pressures and escape the predator-prey cycle by either producing more offspring, produce offspring at decreased sizes at first reproduction when threatened by predators for large prey or large offspring when contacted by a size-limited predator (Lass & Spaak 2003). Inducible defences also hinder developmental processes when threatened by a predator. Several studies have explained this phenomenon: for example, a study on *Ephemerella invaria* larvae showed that when the population sizes of fishes are abundant, the larvae reproduce earlier and have smaller sizes (Dahl & Peckarsky 2003). Furthermore, to escape predation pressure temporarily, organisms develop resting stages as seen for example in dinoflagellates which develop cysts to evade parasites (Toth *et al.* 2004) or zooplankton grazers (Rengefors *et al.* 1998).

Thus, understanding the respective roles of both genetic and phenotypic variation and their interplay with the environment is essential to predict the consequences of adaptation.

I-2 Assessing the genetic basis of local adaptation

Natural populations often show genetic variation at both sequence and gene expression levels. Therefore, assessing their effect on an organism's fitness aids in understanding the genetic basis of local adaptation (MacManes & Eisen 2014).

I-2.1 Expression-based divergence

One of the most commonly used approaches to study the evolution of gene regulation is to quantify and compare gene expression patterns across populations or species with the goal of understanding genetically regulated differences. Gene expression patterns play an important role in the evolution of complex traits and contribute to species-specific adaptations (Wray *et al.* 2003). Populations are often subjected to biotic or abiotic stressors and several studies have examined and reported their transcriptional response. For example, adaptive differences in gene expression have been reported in several single-gene studies (reviewed in Wray *et al.* 2003), but after the advent of high-throughput sequencing technologies, this hypothesis has been tested rigorously because they can assay many loci at once (Whitehead & Crawford 2006a). Furthermore, substantial variation in gene expression for a specific environmental stressor has been widely agreed to have a genetic basis of adaptation as seen in yeast (e.g. Brem *et al.* 2002), *Drosophila* (e.g. Jin *et al.* 2001), mice (e.g. Pritchard *et al.* 2001), and humans (e.g. Sharma *et al.* 2005). Moreover, identifying the underlying differences in expression profiles can help in the prediction of gene functions (Jansen & Nap 2001; Kesari et al. 2012; Rockman 2008). Thus, gene expression patterns play an important role to learn about adaptive processes that drive evolution.

I-2.2 Sequence-based divergence

Identifying loci under potential selection is detected based on F-statistics (Rousset 1997) or principal component analysis (PCA)-based approaches (Jolliffe & Cadima 2016). Fixation-indexes (or F_{ST}) based statistics is a widely used method to identify genetic loci that reveal genetic diversity between populations. Inferring the genetic ancestry differences between individuals from different populations i.e., the population structure, can be visualized with the help of principal component analysis (PCA) plots. To distinguish

DNA sequences that evolve neutrally from selection within populations, measures such as Tajima's D are widely used. A negative value of Tajima's D might indicate population size expansion after bottleneck or selective sweep and purifying selection. A positive D value may indicate population size reduction and balancing selection (Tajima 1989). However, identifying loci under potentially under adaptation comes with its own challenges. First, only a few fixed differences are expected over short evolutionary times among populations. Second, variation among populations reflects not only patterns of selection but also population size, gene flow, genetic drift and demographic histories of populations (Tajima 1989). Thus, genetic differentiation within and among populations is severely affected by population structure, demography and incomplete lineage sorting, which are all randomly distributed across the genome (Rausher & Delph 2015). Therefore, identifying the genetic features underlying local adaptation is critical to understand the process of natural selection.

I-3 Linking genotypes and phenotypes

Since the advent of next-generation sequencing technologies, population genetics research has undergone revolutionary changes (Radwan & Babik 2012). With affordable sequencing costs, data collection from genomes of several individuals has become a reality (Fumagalli *et al.* 2014). From exploring a limited number of genes (for instance, mtDNA, microsatellites or very few nuclear markers) to exploring genome-wide data, the field of genomics has undergone revolutionary developments in the last decades (Veeramah & Hammer 2014). This has led to inferring the evolutionary history and identifying genetic loci associated with a trait in any species at very high resolutions.

One of the major objectives in the field population genetics is to identify adaptive loci associated with phenotypes (Pardo-Diaz *et al.* 2015). Although enormous advances have been made to understand the individual contributions of genotype and phenotype, dissecting the link between them remains a challenge. One of the most common methods to identify genetic architecture of the adaptive trait when phenotypic responses are known is genome-wide association studies (GWAS). GWAS approaches are used to identify variants associated with a phenotypic trait of interest by exploiting linkage disequilibrium measures across thousands of markers (Stranger *et al.* 2011). GWAS has been traditionally

used in several human studies where the researchers pinpoint a causal variant associated to disease phenotype. However, GWAS used in ecological contexts is rare and the number of loci identified using GWAS and identification of causal variants is challenging. Furthermore, most of the studies have only been able to identify a small fraction of variants associated to the traits of interest (Ambrosone 2007). Therefore, utilizing GWAS along with genome-wide scan approaches might help to infer local adaptation process more precisely.

I-4 Daphnia *as a model organism*

Daphnia is a small planktonic crustacean belonging to the order Cladocera and feed primarily on algae. Within the genus *Daphnia, D. galeata* belongs to the *Daphnia longispina* species complex. Closely related species of *D. galeata* in this complex are *D. cucullata* and *D. longispina* and frequent hybridizations take place between them (e.g. Gießler 2001; Schwenk 1993; Schwenk *et al.* 2000; Yin *et al.* 2010). Two subspecies of *D. galeata* are present: *D. galeata galeata* and *D. galeata mendotae*. *D. g. galeata* is native to Eurasia and the *D. g. mendotae* dominate the American continent. *D. galeata* is detected in warm and eutrophic lakes (e.g. Keller *et al.* 2008; Yin *et al.* 2010) and in alpine habitats with lower trophic levels (e.g. Petrusek *et al.* 2007).

Daphnia is an important keystone species found in many freshwater habitats and hold a key position in the aquatic food web. They are a primary consumer of phytoplankton and are an important food source for the secondary consumers such as fish and insects, thereby defining it as a strong ecological interactor (Figure 2; Miner *et al.* 2012). Several characteristics of *Daphnia* make them an invaluable species for experimental genetic studies ranging from studies on population structure and ecology to toxicological studies (Shaw *et al.* 2008). Among them, the most prominent is their complex life cycle. Daphnids are cyclical parthenogens i.e., they are capable of reproducing both asexually and sexually (Hebert 1978). Because of their clonal reproduction, their genetic background allows for the maintenance of permanent intact genotypes (Hebert & Ward 1972), thereby providing an effective defined genetic background for comparisons of various stressor treatments. Sexual reproduction in *Daphnia* is induced by environmental cues, which involves the production of ephippia (Olmstead & LeBlanc 2003). Ephippia are

thick chitinous shells containing up to two long-lived, dormant propagules and accumulate in lake sediments (Burge *et al.* 2018). Lake sedimentation provides a unique opportunity for archiving ephippia and stores information on the environmental parameters such as eutrophication, oligotrophication and pollution levels captured in the sediment cores (Shaw *et al.* 2008). This process helps to preserve the genetic information of *Daphnia* populations, thereby aiding in the investigation of past population history to their modern descendants (Hairston Jr *et al.* 1999). Thus, ephippia can be used for resurrecting old populations in the laboratory and study their phenotypic responses to various environmental stressors (Kerfoot *et al.* 1999) and also infer environmental changes that are responsible for current genomic patterns (Orsini *et al.* 2013a). Resurrection methods have been used in several studies. For example, a study by Frisch *et al.* (2014) assessed the evolutionary responses to eutrophication in *D. pulicaria* using resurrected samples that were at least 700 years old. All the above-mentioned parameters along with existing genomic and ecological resources make *Daphnia* an interesting study species in ecology and evolution.

I-4.1 Daphnia *Genomics*

Daphnids are well established ecotoxicological model organisms and recent advances in sequencing technologies made the availability of three *Daphnia* species genomes viz., *D. pulex* (Colbourne *et al.* 2011), *D. magna* and *D. similoides* and one transcriptome, viz., *D. galeata* (Huylmans *et al.* 2016), a reality. In order to curate, archive and share functional and genomic data, the wFLEABASE (Colbourne *et al.* 2005) was developed by the *Daphnia* Genomics Consortium (DGC) and currently serves as a hub for obtaining gen-/transcript-omic data for the emerging model organism, *Daphnia*.

There are several reasons why genomics approaches are an invaluable tool for *Daphnia*. First, although resurrection ecology is a powerful method to explore the evolutionary dynamics of natural populations, this method has its own limitations i.e., ephippia remain viable and have a >75% hatching success from sediments as old as 20 years and further reduced (<1%) when hatched from sediments dating hundreds of years (Burge *et al.* 2018). Therefore, existing studies focussed on resurrected *Daphnia* that are <70 years old (e.g. Cousyn *et al.* 2001). However, a few exceptions exist such as Frisch *et al.*

(2014) which have used resurrected *Daphnia* samples that are ~700 years old. To overcome this constraint, extracting DNA that the ephippia contain can broaden this timeline to hundreds or even thousands of years, even when these resting eggs are no longer viable. Once ecologically relevant genes are identified, information from molecular signatures of trait changes can be used to infer evolutionary patterns (Joachim *et al.* 2007).

Second, an important aspect of understanding the mechanism of adaptation is to identify candidate genes linked to specific traits and annotating their function. Once the gene function is annotated and its interaction with a specific environmental perturbation is identified, gene homologs present in other species and their relevance to ecological factors can be inferred. Although a genome is a necessity to enable the above-mentioned approaches, the ecological role of *Daphnia* along with its key assets for experiments will be central to enhance our understanding of ecological genomics. Therefore, the extensive knowledge of *Daphnia* ecology and evolution that has been accumulated in the past century should be used for genome investigations (Miner *et al.* 2012).

I-4.2 Local adaptation in Daphnia

For organisms with high dispersal ability such as *Daphnia*, high levels of gene flow should reduce genetic variation that results from selection and drift (Allen *et al.* 2010). For example, higher rates of dispersal may lead to homogenization of gene pools by breaking down of adaptive differentiation among populations (Fuller *et al.* 1996; Gilbert-Horvath *et al.* 2006; Roslin 2001). On the other hand, low rates of dispersal results in increased population genetic differentiation, where selection and drift are less affected (Ehrlich & Raven 1969; Slatkin 1985). *Daphnia* disperses mainly through (a) natural mechanisms such as transportation through water birds, floating in sea ice, flooding and wind; (b) human-mediated mechanisms such as fish stocking and construction of canals and pipelines using equipment (reviewed in Havel 2004). However, despite their high dispersal rates, several studies have shown substantial genetic variation among populations and rapid colonization of new habitats (Allen *et al.* 2010; Cohen & Shurin 2003; Johnson *et al.* 2008; Louette & De Meester 2005). *Daphnia* often show strong population genetic

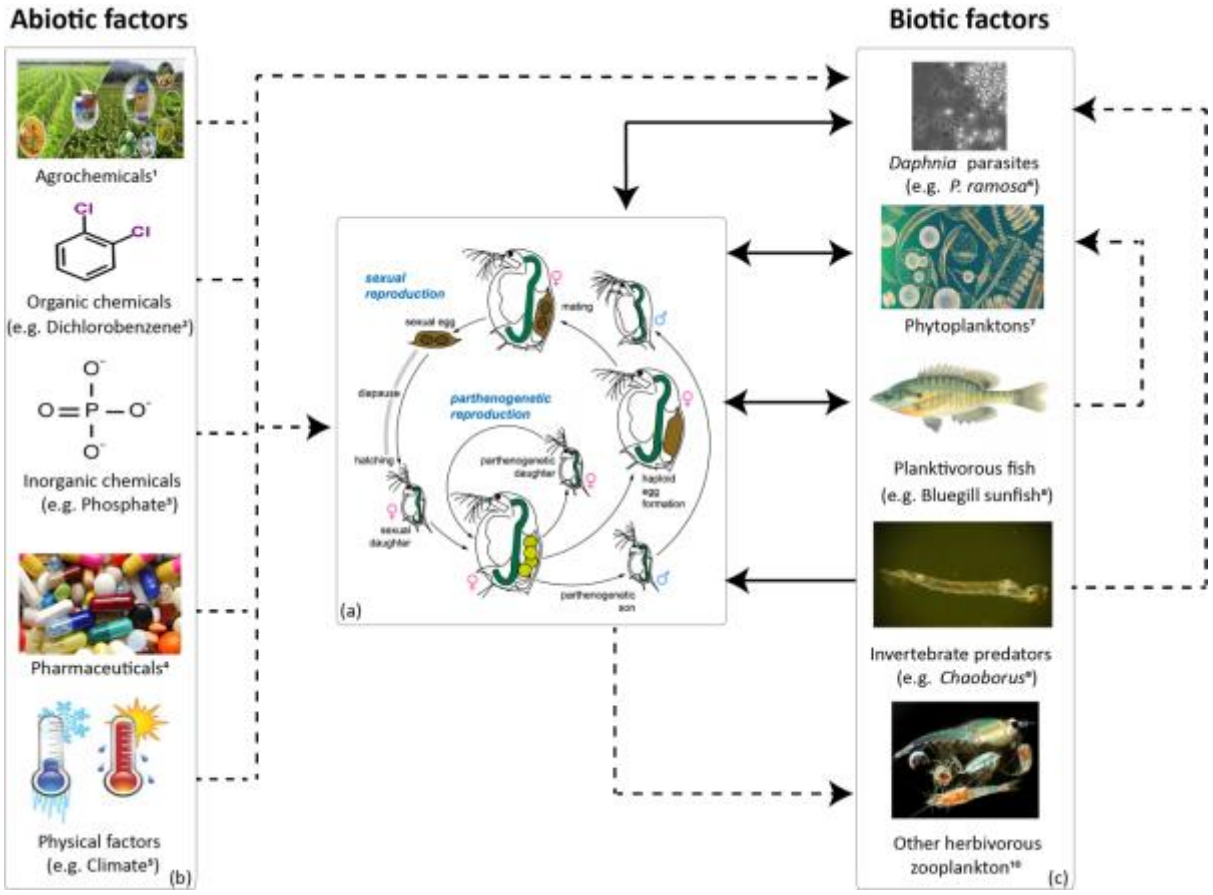


Figure 2: Life cycle of Daphnia and its ecological interactors.
(a) Life cycle of *Daphnia* (adapted from Ebert 2005) and its interactions with (b) abiotic and (c) biotic factors in a lake ecosystem (adapted from Miner et al. 2012). Solid lines represent direct interactions and dashed lines indicate indirect interactions; direction (uni-or bi-directional) of the arrows is based on literature evidence as reviewed by Miner et al. (2012). Image sources can be found in "Appendix for Introduction" section.

differentiation even at small geographical (i.e., less than 1 km) scales (Allen *et al.* 2010; Boileau *et al.* 1992; Haag *et al.* 2006; Hamrova *et al.* 2011; Spitze 1993; Thielsch *et al.* 2009; Vanoverbeke & De Meester 1997a; Yin *et al.* 2010). In addition to this, the monopolization effect, a concept used in several freshwater invertebrate studies (De Meester *et al.* 2002; Louette *et al.* 2007; Munoz *et al.* 2008; Ortells *et al.* 2013) might reinforce population structure resulting from initial colonization events. In habitats recently colonized (less than 5 years) by *Daphnia*, founder effects lead to genetic differentiation among populations that do not erode immediately despite ongoing dispersal (Ortells *et al.* 2013). Thus, a combination of several factors such as gene flow, rates of dispersal, colonization patterns, founder events or bottleneck events, selection can all play an important role in the evolutionary adaptation of *Daphnia* populations.

I-4.3 Inducible defences in Daphnia

In *Daphnia*, inducible defences are mediated by chemicals released by the predator, called kairomones. This chemical messenger benefits the receiver rather than being advantageous for the predator (sender) itself (Brown *et al.* 1970). For example, to locate food or sense other predators, the receiver can use kairomones (Pohnert *et al.* 2007). Although kairomones are ubiquitous and play a major ecological role, the structure of any of the existing kairomones is not characterized or unidentified (Laforsch & Tollrian 2009; Silberbush *et al.* 2010). However, one exception is the characterization of aliphatic sulphates as *Daphnia*-kairomones, which play a prominent role in colony formation in green algae (Yasumoto *et al.* 2008; Yasumoto *et al.* 2006). In addition to kairomones, several factors such as alarm cues by injured conspecifics (Laforsch *et al.* 2006; Pijanowska 1997; Stabell *et al.* 2003) or physical cues such as predator borne turbulences (Laforsch & Tollrian 2004a) are also responsible for inducible defences in *Daphnia*.

A variety of life-history defences are expressed in response to predator kairomones, leading to changes in important growth-related traits such as size, age at maturity (e.g. Stibor & Lüning 1994). When *Daphnia* are exposed to fish, it decreases their size and age at first reproduction as seen in empirical and theoretical predictions on *D. galeata* (Macháček 1995; Stibor 1992), *D. hyalina* (Stibor & Lüning 1994) and *D. magna* (e.g. Lampert 1993; Sakwińska 2008). On the contrary, an increased size and age at first reproduction is observed when *Daphnia* are exposed to invertebrate predators such as *Chaoborus*, as seen for example in *D. hyalina* (Stibor & Lüning 1994) and *D. pulex* (e.g. Boeing *et al.* 2006; Tollrian 1995). Similarly, contrasting patterns of size and age at reproduction i.e., increased size and early maturation has been shown in *D. magna* when exposed to cyclopoids (Pijanowska & Kowalczewski 1997). Several other shifts in life-history traits such as decreased size at first reproduction and increased number of offspring are observed when daphniids are exposed to fish such as *D. pulex* (e.g. Dodson 1989), *D. hyalina* (e.g. Stibor 1992) and *D. magna* (e.g. Boersma *et al.* 1998), while invertebrate predators, such as *Chaoborus* induce contrary reactions, e.g. in *D. pulex* (Black & Dodson 1990) and *D. hyalina* (Stibor & Lüning 1994).

Very few studies (Bento *et al.* 2017; Henning-Lucass *et al.* 2016; Herrmann *et al.* 2017a; Schwerin *et al.* 2009) have assessed the genotype-phenotype associations in

Daphnia under different environments. A study by Routtu *et al*. (2014) and De Meester (1991) have identified eQTLs that contribute to variation in one trait namely, phototactic behaviour in the presence of fish kairomones. However, none of them have explored the genetic basis for any of the reproduction-related phenotypic traits in the absence and presence of predator kairomones. Therefore, understanding the underlying genetic basis of the life-history traits in a fish kairomone environment is necessary to evaluate and predict the consequences of inducible defences in *Daphnia*.

Thesis Outline

Aims of the study

In this thesis, I aim to assess the transcriptomic basis of local adaptation in European *D. galeata* populations and infer their genotype-phenotype relationships. Furthermore, I want to improve the existing functional annotation *D. galeata* transcripts, thereby identifying their ecological roles more precisely. Therefore, I address the following questions:

1. What is the role of natural selection and genetic drift on gene expression and sequence variation within and between four lake populations of *D. galeata*?
2. What is the functional role of the *D. galeata* transcripts?
3. What conclusions can be drawn from the genotype-phenotype relationships of *D. galeata* in the absence and presence of a predator kairomone?

Chapter Overview

In **Chapter 1**, I focus on understanding transcriptomic variation at the sequence and regulatory level in four lake populations of European *D. galeata*. To this aim, I used *D. galeata* transcriptome data to identify differentially expressed transcripts and performed a sequence analysis to identify candidates for local adaptation. This analysis helped to reveal contrasting patterns of divergence at both sequence and gene expression levels. Furthermore, to understand the role of each transcript, I annotated their function using comparative genomics approaches. I observed that most of the *D. galeata* transcripts functionally annotated were "hypothetical" or of "unknown" function. Hence, looking deeper into the function of the hypothetical transcripts would help in pinpointing the ecological roles of *Daphnia* genes/transcripts.

Therefore, in **Chapter 2** I focused on enriching the functional annotation of *Daphnia* transcripts. To overcome this challenge, I screened the literature for gene expression studies in *Daphnia*, manually mining their expression profiles and responses to various experimentally validated environmental stressors. I used a homology approach to link *D. galeata* transcripts to *Daphnia* specific genes whose expression patterns are shown to be affected by one or more stressors as identified in the literature. This approach helped me

to identify transcript-specific stressors in *D. galeata*, thereby providing insights into their ecological roles. The literature data obtained in this meta-analysis was developed into a database, which is publicly available to help fellow researchers, thereby providing a tool that would help in the interpretation of future ecotoxicological studies.

After assessing the transcriptomic and functional profiles of *D. galeata* extensively, it is important to assess their genotype-phenotype relationships, as it allows drawing conclusions about the extent of a SNPs contribution to each of the life-history traits. Therefore, in **Chapter 3**, I focused on understanding the genotype-phenotype relationships using SNP data obtained in **Chapter 1** and life-history phenotypic traits in the absence and presence of predator fish kairomones obtained using the same 24 clonal lines of *D. galeata* from another experiment. Furthermore, a weighted gene co-expression network analysis helped to correlate the phenotypic traits and gene expression profiles thereby identifying clusters of transcripts likely to be involved in reproduction-related traits.

Chapter 1: Contrasting patterns of divergence at the regulatory and sequence level in European *Daphnia galeata* natural populations

Suda Parimala Ravindran, Maike Herrmann and Mathilde Cordellier

Abstract

Understanding the genetic basis of local adaptation has long been a focus of evolutionary biology. Recently there has been increased interest in deciphering the evolutionary role of *Daphnia*'s plasticity and the molecular mechanisms of local adaptation. Using transcriptome data, we assessed the differences in gene expression profiles and sequences in four European *Daphnia galeata* populations. In total, ~33% of 32,903 transcripts were differentially expressed between populations. Among 10,280 differentially expressed transcripts, 5,209 transcripts deviated from neutral expectations and their population-specific expression pattern is likely the result of local adaptation processes. Furthermore, a SNP analysis allowed inferring population structure and distribution of genetic variation. The population divergence at the sequence-level was comparatively higher than the gene expression level by several orders of magnitude consistent with strong founder effects and lack of gene flow between populations. Using sequence homology, the candidate transcripts were annotated using a comparative genomics approach. Additionally, we also performed a weighted gene co-expression analysis to identify population-specific regulatory patterns of transcripts in *D. galeata*. Thus, we identified candidate transcriptomic regions for local adaptation in this key species of aquatic ecosystems in the absence of any laboratory induced stressor.

Keywords: constitutive gene expression, RNA-seq, molecular phenotype, population transcriptomics, DRIFTSEL, WGCNA

This version of the chapter has been submitted to *Ecology and Evolution*

Introduction

Natural genetic variation shapes divergence in phenotypic traits and is an important resource for evolutionary processes (Oleksiak *et al.* 2002). Populations respond to environmental variation by genetically adapting to their environments (Hereford 2009; Kawecki & Ebert 2004; Savolainen *et al.* 2013), often showing variations at both gene expression and sequence level across the geographic range of a species. One of the fundamental goals of research in the field of molecular evolution is to resolve the evolutionary processes driving the rise and maintenance of expression and sequence polymorphisms behind this variation. Revealing their effect on an organism's fitness thereby aids to understand the genetic basis of local adaptation (MacManes & Eisen 2014). Gene expression patterns link genotypes and phenotypes, sometimes called a "molecular phenotype", and as such is an important component in local adaptation processes (Lopez-Maury *et al.* 2008). --Several studies have reported the testing of different populations exposed to different treatments and examining their transcriptional response, for example in springtails (*Folsomia* (De Boer *et al.* 2013) and *Orchesella* (Roelofs *et al.* 2009)), oyster (*Crassostrea virginica*; Chapman *et al.* 2011; Chapman *et al.* 2009), sparrows (*Zonotrichia capensis*; Cheviron *et al.* 2008), flounder (*Platichthys flesus*; Larsen *et al.* 2008), and seagrass *(Zostera marina*; Jueterbock *et al.* 2016; Reusch *et al.* 2008), thereby identifying candidate genes involved in local adaptation. Gene expression variation can be highly heritable (Brem & Kruglyak 2005; Schadt *et al.* 2003; Whitehead & Crawford 2006b). Moreover, constitutive gene expression patterns also differ within- and among-natural populations (e.g., Roberge *et al.* 2007; Whitehead & Crawford 2006a), strongly suggesting that standing variation in constitutive gene expression is shaped by local adaptation. Natural selection acts immediately on newly arisen variation (in contrast to adaptation observed from standing genetic variation) as there are neutral and slightly deleterious variations preserved in a population, which may become beneficial upon changes in selection regimes (Barrett & Schluter 2008). After a sudden change of environment, standing variation can contribute to fast adaptation (Feulner *et al.* 2013; Kitano *et al.* 2008). Identifying allelic/genetic variants underlying differences in expression profiles can be helpful in hypothesizing gene functions (Jansen & Nap 2001; Kesari *et al.* 2012; Rockman 2008). Although prior knowledge of the specific loci is not a prerequisite to

learn about adaptive processes in most cases, identification of genetic features underlying local adaptation is critical in answering fundamental questions about natural selection (Rausher & Delph 2015).

Genetic variation within and among populations is strongly influenced by their colonization history, and the demographic changes following the primary establishment of a population. Population sizes may vary after colonization across the species based on environmental factors and further colonization (Böndel *et al.* 2015). Colonization events depend on dispersal ability, and dispersal rates strongly differ from gene flow estimates in several species (De Meester *et al.* 2002). This is particularly evident in freshwater zooplankton species, where several studies suggest a high potential for dispersal when populations rapidly colonize new habitats and spread invasively (Havel *et al.* 2000; Louette & De Meester 2004; Mergeay *et al.* 2008). However, genetic studies show that the observed rate of gene flow is much lower than would be expected in organisms with high dispersal potential (Boileau *et al.* 1992; De Meester *et al.* 2002; Thielsch *et al.* 2009). This ambiguity between dispersal potential and rate of gene flow can be explained by founder effects (Boileau *et al.* 1992) complemented by local adaptation; resulting in monopolization of resources by local populations (De Meester *et al.* 2002). This process leads to the impression that population genetic variation correlates with the colonization patterns (Orsini *et al.* 2013b).

Amongst freshwater zooplankton species, the water flea *Daphnia* (Figure 1) is the best studied and has been an important model for ecology, population genetics, evolutionary biology, and toxicology (Ebert 2005). This genus belongs to the order Cladocera and has attracted scientific interest since the 17th century (Desmarais 1997). It inhabits most types of freshwater habitats and includes more than 100 known species of freshwater plankton organisms (Ebert 2005). *Daphnia* make an interesting subject of investigation in comparative functional genomics (Eads *et al.* 2008). Apart from the fact that *Daphnia* species have an appropriate size for being used in laboratory cultures, they are easy to cultivate and have short generation times. Because of their clonal mode of reproduction, *Daphnia* are highly suited for quantitative genetic studies, which can enhance our understanding of their evolutionary ecology.

Figure 3: Waterfleas, *Daphnia galeata*, Photo: Mathilde Cordellier

Genetic variation has been reported for numerous traits in *Daphnia,* such as life history traits (e.g., Henning-Lucass *et al.* 2016), vertical migration (e.g., Haupt *et al.* 2009), fish escape behavior (e.g., Pietrzak *et al.* 2015), resistance against parasites (e.g., Routtu & Ebert 2015) and immune response (e.g., Garbutt *et al.* 2014). Furthermore, it was shown that responses to many chemical stressors such as phosphorus (Roy Chowdhury *et al.* 2015; Roy Chowdhury *et al.* 2014), copper (Poynton *et al.* 2008), cadmium (Soetaert *et al.* 2007) and pharmaceutical products like ibuprofen (Hayashi *et al.* 2008; Heckmann *et al.* 2007) have a genetic basis as well. Within- and between-population comparisons in *Daphnia* have been conducted extensively using varied environmental perturbations providing evidence for local adaptation (for e.g., Barata *et al.* 2002; Declerck *et al.* 2001; Ebert *et al.* 1998; Spitze 1993). Although various aspects like phylogeography, functional morphology, physiology and life history evolution have been in the limelight of *Daphnia* research for several decades (Eads *et al.* 2008), *Daphnia* genomics investigations have begun only in the last decade with the availability of the *Daphnia pulex* genome (Colbourne *et al.* 2011). A considerable number of studies (for e.g.: Bento *et al.* 2017; Miner *et al.* 2012; Orsini *et al.* 2016; Yampolsky *et al.* 2014) on biotic and abiotic factors have been carried out showing how *Daphnia* respond to environmental perturbations by changes in gene expression. However, little is known about the intra-specific variability at the gene expression level in

Daphnia, since the above-mentioned studies focused on stressor driven responses using a limited number of genotypes.

To sum up, elucidating the mechanisms by which natural selection acts on gene expression evolution remains a challenge (e.g.: Fraser 2011; Romero *et al.* 2012). Unraveling the relative consequences of drift versus natural selection on gene expression profiles plays an important role in understanding species divergence and local adaptation. The studies listed above provided evidence for gene expression variation correlated with many environmental factors in *Daphnia*. However, knowledge about the variation in constitutive gene expression structure within and among population is lacking.

In the present study on *Daphnia galeata,* sampled from four different lakes in Europe, we conducted a large-scale RNA-seq study in the absence of any laboratory induced environmental stressor. Using transcriptome data, we quantified the constitutive expression profiles and performed a sequence analysis of the four populations. We addressed the following questions: (i) Are there differences in gene expression profiles between the four populations? (ii) How is the observed variation explained by the different levels of organization, i.e., genotype and population? (iii) Do the observed differences in expression profiles result from genetic drift or selection? (iv) What is the role of genetic drift and/or natural selection in shaping sequence variation? (v) What are the functional roles of the transcripts?

Our study brought contrasting patterns of divergence at the regulatory and sequence level into light. While no population specific gene expression patterns were found for majority of the analyzed transcripts, divergence patterns at the sequence level hinted at strong influences of founder effects, bottleneck events and divergent selection. Further, our gene co-expression network analysis showed conserved patterns while assessing the population–specific networks and supported our observations at the regulatory level. We were able to identify candidate transcripts for local adaptation using combined approaches. Further comparative genomics analyses are needed to complement our preliminary functional annotations of these candidate transcripts to identify the ecological drivers behind the observed patterns of adaptation.

Methods

Sampling and RNA collection

A set of *D. galeata* resting stages (ephippia) was collected from the sediment of four lakes: Jordán Reservoir (hereafter, Pop.J) in Czech Republic, Müggelsee in Germany (hereafter, Pop.M), Lake Constance (hereafter, Pop.LC) at the border between Germany, Switzerland and Austria, and Greifensee (hereafter, Pop.G) in Switzerland. These ephippia were hatched under laboratory conditions (see Henning-Lucass *et al.* 2016 for hatching conditions) and the hatchlings were used to establish clonal lines in a laboratory setting. The species identity was checked by sequencing a fragment of the *12S* mitochondrial locus and 10 microsatellite markers (Multiplex 2 comprising the loci *Dgm109, Dp196, Dp281, Dp512, SwiD1, SwiD10, SwiD12, SwiD14, SwiD15, SwiD2*), following protocols by Taylor *et al.* (1996) and Yin *et al.* (2010) respectively.

Mature females for six clonal lines per lake were placed at equal densities (40 individuals L^{-1}) in semi-artificial medium for a week, during which the juveniles were regularly removed. Gravid females from the equal density beakers were then collected within three days during a time window of a few hours. Twenty to thirty individuals were homogenized in a 1.5 mL centrifuge tube in 1 mL Trizol (Invitrogen, Waltham, MA USA) immediately after removing the water. The Trizol homogenates were kept at -80 °C until further processing.

RNA preparation and sequencing

Total RNA was extracted following a modified phenol/chloroform protocol and further processed using the RNeasy kit (Qiagen, Hilden, Germany). The total RNA was eluted in RNAse free water and the concentration and quality (RNA integrity number and phenol contamination) were checked using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The 72 total RNA samples (4 lakes x 6 genotypes x 3 biological replicates) were sent to the company GATC (Konstanz, Germany) for library preparation and sequencing. Following reverse transcription and cDNA construction using random primers, 50bp single-end (SE) reads were sequenced on an Illumina HiSeq 2000 (San Diego, CA, USA). To avoid block effects and confounding effects in the downstream analysis, we used a completely

randomized design; each library was sequenced on at least two different lanes, on a total of nine lanes. Detailed information can be found in Table S1.

Quality trimming, mapping and read counts

All reads with ambiguous bases (Ns) were removed before trimming. Bases with a phred score below 20 were trimmed at the 3' and 5' ends. Reads shorter than 45 bp after trimming were discarded. All trimming steps were conducted using locally installed version of Galaxy at the Gene Center in Munich, Germany.

Trimmed reads were mapped to the reference *D. galeata* transcriptome (Huylmans *et al.* 2016; available from NCBI: https://www.ncbi.nlm.nih.gov, GenBank ID: HAFN00000000.1) using NextGenMap (Sedlazeck *et al.* 2013) with increased sensitivity (-i 0.8 –kmer-skip 0 -s 0.0). Read counts were obtained from the SAM files using a custom python script (available upon request) and discarding ambiguously mapped reads. The raw count table was analyzed in R (R Development Core Team 2008) using the package DESeq2 (Love *et al.* 2014). Normalization was done with size factor procedure. Standard differential analysis steps of DESeq2 such as estimation of dispersion and negative binomial GLM fitting were applied. The count outliers were automatically detected using Cook's distance, which is a measure of how much the fitted coefficients would change if an individual sample was removed (Cook 1977). Principal Component Analysis (PCA) was performed to visualize the clustering of biological replicates and clonal lines.

To identify the differentially expressed transcripts (DETs) upregulated the most in each population, we used the DESeq2 "contrasts" function. We performed six pairwise comparisons: Pop.G vs Pop.J, Pop.G vs Pop.LC, Pop.G vs Pop.M, Pop.J vs Pop.LC, Pop.J vs Pop.M, Pop.LC vs Pop.M. All *p*-values were adjusted for multiple testing using the Benjamini-Hochberg correction (Benjamini 1995) implemented in DESeq2. To create a list for each population from each comparison, we retained transcripts that had an adjusted *p*-value (p_{adj}) equal to or lower than 0.05 and a fold change (FC) deviating from 0 (depending on the direction of the pairwise comparison), resulting in four lists as follows:

1. Pop.G: G vs. M: FC > 0; G vs. LC: FC > 0; J vs. G; FC < 0
2. Pop.J: J vs. G: FC > 0; J vs. LC: FC > 0; J vs. M; FC > 0
3. Pop.LC: J vs. LC: FC < 0; LC vs. M: FC > 0; G vs. LC; FC < 0

4. Pop.M: G vs. M: FC < 0; J vs. M: FC < 0; LC vs. M; FC < 0

The four lists of DETs obtained above were combined to identify population specific transcripts and Venn diagrams depicting the overlap between the contrasts were created using the VennDiagram package (Chen 2011) in R.

Evaluating the role of natural selection on transcript expression levels: DRIFTSEL

We searched for transcripts for which the identified differential expression could not be explained by phylogenetic distance and genetic drift alone. To identify signals of possible selection, we used the approach of Ovaskainen *et al.* (2011) implemented in the R package DRIFTSEL 2.1.2 (Karhunen *et al.* 2013), considering expression of every single transcript as a trait. To perform this analysis, we made use of the microsatellite data and normalized read count values. Allele frequencies were obtained from microsatellite data collected in a previous study, independently from the species identification step outlined above. Microsatellite data of 30-40 resting eggs also sampled from the same sediment layers the resurrected clonal lines come from was obtained from a study by Herrmann (2017). Briefly, eleven microsatellite loci were analyzed for each clonal line according to the protocol published by Thielsch *et al.* (2009). Primers for all loci were multiplexed and PCR was performed using the Type-it Microsatellite PCR Kit (Qiagen, Hilden, Germany). Alleles were recorded manually and allelic frequencies were calculated with GenAlEx (Peakall & Smouse 2012).

Using microsatellite allelic frequencies, the coancestry coefficients by admixture F model was calculated using "do.all" function implemented in the RAFM package (Karhunen & Ovaskainen 2012). We ran a total of 200,000 iterations with thinning at an interval of 1,000 and discarded the first 1,000 iterations as 'burn-in'. The output was a list which contained samples from the posterior distributions of allele frequencies. Values from the posterior coancestry matrix, 'theta', were used as input for the Metropolis-Hastings (MH) algorithm along with the normalized "mean" read counts of the replicates (i.e. one value per genotype) for each DET as implemented in DRIFTSEL. We ran a total of 5,000 iterations with thinning at 1,000 samples and discarded the first 100 iterations as burn-in. The output of MH algorithm was a matrix of posterior of subpopulation effects (pop.ef), used to estimate the H.test values. The H.test describes whether the population

means correlate with the genetic data more than it would be expected on basis of shared evolutionary history. Large H-values (H-value ≥ 0.95) imply that the populations are more locally adapted than expected by chance.

Intra and inter-population variation

To quantify the respective contributions of the factors "genotype" and "population" to the observed variation in gene expression profiles, we performed a linear mixed model analysis in R on the normalized read counts obtained from DESeq2. We used the slope of 'Genotypes' as a random factor (to account for multiple replicates per genotype per population) and 'Population' and 'Genotype' as fixed factors. To compute pvalues for our model, we used the 'Anova' function in the R package "car" (Weisberg 2011). To correct for multiple testing, p_{adj}-values were calculated for each transcript using the Benjamini-Hochberg procedure.

Variant calling and filtering

The variant calling and filtering steps have already been described in Herrmann *et al.* (2017c). Briefly, the aligned reads from RNA-seq data were merged using samtools (Li *et al.* 2009). GATK (McKenna *et al.* 2010) was used to split exon segments, reassign the mapping qualities (SplitNCigarReads) and indels were aligned (RealignerTargetCreator and IndelRealigner). The HaplotypeCaller (DePristo *et al.* 2011) function was used for the initial variant calls for the realigned reads and samples were jointly genotyped using GATK's GenotypeGVCFs tool. A single vcf file was created and false positive variant calls were filtered with the following criteria: (i) clusterWindowSize = 35; (ii) Quality by depth (QD) < 2.0; (iii) Fisher Strand (FS) > 30.0. This produced a variant dataset with not only biallelic variants but also triallelic variants and indels.

Using the SNPRelate package (Zheng *et al.* 2012) in R/BioConductor, the variant dataset was limited to only biallelic sites for downstream analysis. These were further pruned for linkage disequilibrium considering a threshold of 0.2 (r^2 > 0.2), thereby retaining 393,514 SNPs. A PCA was plotted using the functions in SNPRelate which include calculating the genetic covariance matrix from genotypes, computing the correlation coefficients between sample loadings and genotypes for each SNP, calculating SNP

eigenvectors (loadings) and estimating the sample loadings of a new dataset from specified SNP eigenvectors.

Neutrality statistics

To obtain alignments of transcript sequences, SNP calling datasets were filtered as described above. Beagle 4.1 (Browning & Browning 2007) was used to phase SNP calling data and a python script (available upon request) was used to parse the phased vcf file to sample sequences in fasta format. After phasing, we obtained 13006 transcripts containing SNPs and the sequences were input in R. A multiple sequence alignment and Tajima's D statistics (with *p*-values) were obtained population-wise for each transcript using the pegas package (Paradis 2010) in R.

Results from LOSITAN (Antao *et al.* 2008) outlier tests were obtained from Herrmann *et al.* (2017c) to identify loci under selection (see Table S4). Briefly, under the infinite allele model, 500,000 simulations were conducted with a confidence interval of 0.99, false discovery rate of 0.1, attempted F_{ST} of 0.182, subsample size of 12 (as computed by LOSITAN) and simulated F_{ST} of 0.181. For more details on how LOSITAN analysis was performed, see Herrmann *et al.* (2017c).

Inbreeding coefficient and mutation frequencies

The inbreeding coefficient for the final SNP dataset were calculated with the --het function in VCFtools, (Danecek *et al.* 2011). The ratio between the expected heterozygosity (H_E) and observed heterozygosity (H_O) was calculated based on available SNP information and plots were created using ggplot2 (Wickham 2009) in R.

Sequence vs. regulatory variation

To visualize the proportion of transcripts responsible for local adaptation at regulatory and sequence level, we consolidated the list of transcripts from various analyses as performed above and represented it with an alluvial diagram (http://rawgraphs.io/). In an alluvial diagram, each black rectangle is called a 'node', the colored regions linking the nodes are called 'flows' and the vertical group of nodes are

called 'steps'. In our analyses, we had four steps: DESeq2, DRIFTSEL, LOSITAN and Tajima's D.

Annotation and functional analysis

To functionally annotate the *D. galeata* transcripts, a local sequence alignment using blastn (Altschul *et al.* 1990) against the nr database (downloaded Feb. 2015 via ftp://ftp.ncbi.nlm.nih.gov/blast/db/) was performed. Hits with an evalue ≤ 0.1 and identity ≥ 50% were considered. Additionally, protein domain annotations and orthoMCL (Li *et al.* 2003) results were obtained from Huylmans *et al.* (2016) . Briefly, a search was made for all three *Daphnia* species (*D. pulex, D. magna* and *D. galeata*) using PfamScan (version 1.5) to look into the Pfam A database (version 27.0; Finn *et al.* 2014) together with hmmer3 (version 3.1b; Mistry *et al.* 2013). In order to identify orthologs and be able to compare it to other arthropod species, orthoMCL was used to cluster the amino acid sequences of *D galeata*, *D. pulex* (version JGI060905; Colbourne *et al.* 2011), *D. magna* (version 7; *Daphnia* Genomics Consortium 2015), as well as *Drosophila melanogaster* (version 5.56; St Pierre *et al.* 2014) and *Nasonia vitripennis* (version 1.2; Werren *et al.* 2010) into orthologous groups and determine the inparalogs. Pie charts representing the number of hits obtained for all transcripts and DETs were created using the plotrix package (Lemon 2006) in R.

We classified the orthoMCL clusters into the following categories:

(a) Clusters that contain only *D. galeata*-specific transcripts
(b) Clusters that are shared between *D. galeata* and *D. pulex*
(c) Clusters that are shared between *D. galeata* and *D. magna*
(d) Clusters that are shared between *D. galeata*, *D. pulex* and *D. magna* (*Daphnia*-specific)
(e) Clusters that are shared between *D. galeata* and other arthropods (*D. melanogaster* and *N. vitripennis*)
(f) Clusters that are shared among all five analyzed species (*Daphnia* and both insects).

Inparalogs and misassemblies

To assess whether *D. galeata* DETs in an orthologous group are "inparalogs", isoforms or the result of misassembly, we computed the pairwise sequence divergence for those orthoMCL clusters containing DETs from at least two different populations. Since each significantly differentially expressed transcript was assigned as a DET only to the population in which it was upregulated the most, clusters containing more than one DET most likely contained DETs from different populations. Based on the number of populations within their orthoMCL cluster, the DETs were classified into the categories: "1Pop", "2Pop", "3Pop" and "4Pop", and unclustered DETs were categorized as "no-cluster DETs". 'No-cluster DETs' and '1Pop' DETs were excluded from further analysis. In total, there were 716 orthoMCL clusters that contained DETs from at least two populations. Pairwise alignments of the amino acid sequences in each orthologous group were performed using the iterative refinement method incorporating local pairwise alignment information (L-INS-i) in MAFFT (Katoh *et al.* 2002). We then used EMBOSS tranalign (Rice *et al.* 2000) to generate alignments of nucleic acid coding regions translated from aligned protein sequences. Pairwise genetic divergence was computed with 'dist.dna' function implemented in the ape package (Paradis *et al.* 2004) in R, using the Kimura-2-parameters model with gamma correction. We used an arbitrary cut-off value of 2 to distinguish inparalogs from misassembled sequences.

Gene Ontology (GO) enrichment analysis

DETs with a H.value ≥ 0.95 (DRIFTSEL result) and transcripts with a nonzero D value in each of the four populations (Tajima's D result) were analyzed with „topGO" (Alexa & Rahnenfuhrer 2016) in R, using a custom GO annotation for *D. galeata*. GO terms enriched in the transcripts of interest in each population from each analysis (DRIFTSEL and Tajima's D) were identified using the 'weight01' algorithm for all three ontologies, namely: molecular function, cellular component and biological processes. We used a Fisher test and those GO terms with a classicFisher value ≤ 0.05 were considered to be enriched for each ontology in each population. A multiple testing procedure was not applied as the *p*-values returned by the 'weight01' algorithm are interpreted to be corrected and might exclude "true" annotations (Alexa & Rahnenfuhrer 2016).

Weighted Gene Co-expression network analysis

To gain insights into the population-specific regulatory patterns of transcripts in *D. galeata*, we performed a weighted gene co-expression network analysis with WGCNA (Langfelder & Horvath 2008) using the variance stabilized normalized read counts obtained in DESeq2 analysis. Transcripts and samples that had lower expression values were excluded from every population using the 'goodSampleGenes' function in WGCNA and used for downstream analysis. In total, 32375 transcripts were used for the construction of gene co-expression networks. To identify population-specific co-expression modules (i.e., clusters of highly correlated transcripts), a network was first built using the full dataset (i.e., with samples and transcripts from all populations) and one network for each population using expression values specific to all genotype and biological replicates. The population specific network was compared to the reference network and an adjacency matrix was calculated. Clusters were identified using the WGCNA Topological Overlap Matrices (TOM). For every transcript and module detected automatically, WGCNA assigns a color based on the module membership (MM) value. An MM value is a measure of module membership which is obtained by correlating its gene expression profiles with module eigengene (i.e., the first principal component of a given module). For example, if a transcript has an MMred value close to ±1, the transcript is assigned to the red module (Langfelder & Horvath 2008). Each module is assigned a color based on the module size: 'turquoise' denotes the largest module, blue next, followed by brown, green, yellow and so on. The color 'grey' is reserved for unassigned transcripts (Langfelder & Horvath 2008). Similarly, the module 'gold' consists of 1000 randomly selected transcripts that represent a sample of the whole network and statistical measures have no meaning for this module (Langfelder & Horvath 2008).

After obtaining the module definitions from each comparison, we assessed how well our modules in the reference network are preserved in the population specific networks using the 'modulePreservation' function, which outputs a single Z-score summary. The higher the Z-score, the more preserved a module is between the reference and population-specific network. A module was deemed to be preserved if the Z-score value was above 10, an arbitrary value deemed suitable by Langfelder *et al.* (2011).

Results

Sequencing results and mapping statistics

The dataset used for this study has been described in a previous publication by Herrmann *et al.* (2017c). Between 14 and 30 million reads were obtained for each of the 72 libraries. On average, 95.9% of the data were retained after quality control, and a mean 88.8% were mapped to the reference transcriptome. No mapping bias was observed i.e., very similar results were obtained for all genotypes. All quality and mapping metrics are available on Dryad (Herrmann *et al.* 2017b) and the raw data and experimental setup have been submitted to the ArrayExpress platform (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6144). Raw reads are also available on the European Nucleotide Archive (Study ERP105101; https://www.ebi.ac.uk/ena/data/view/PRJEB23352).

Differential expression

The intraspecific variation in transcript expression in the four populations was visualized from a read counts matrix of the 32903 transcripts using PCA (Figure 4a). A large proportion of the observed variance (19%) is explained by the first principal component (PC1). PC2 and PC3 explained 12% and 10% of the total variance, respectively. Clear population clustering is evident along PC2 for Pop.M and in Pop.J except for two genotypes (J2.1 and J2.4). However, genotypes from Pop.G and Pop.LC belong to overlapping clusters (Figure 4a and Supporting Information Appendix S1). No evident clustering according to experimental parameters (i.e. culture conditions, harvesting, RNA extraction batches) were visible on the PCA.

After conducting pairwise contrast analyses with DESeq2, we identified transcripts exclusively upregulated for each population when compared to all others ($p_{adj} \leq 0.05$; thereafter DETs). In total, 10,820 of 32,903 transcripts (~33%) showed significant expression differences in pairwise comparisons between populations. Of all ~33000 transcripts, 9.6%, 8.1%, 7.2% and 7.8% were population specific DETs for the populations Pop.G, Pop.J, Pop.LC and Pop.M, respectively (Figure 4b).

Role of natural selection on transcript expression levels

The DRIFTSEL multivariate approach was used to identify transcripts for which the observed differential expression could not be explained by phylogenetic distance and genetic drift alone; the alternative explanation being that the observed divergence would be attributable to selection and therefore possibly to local adaptation events. In total, 48% of 10820 differentially expressed transcripts showed greater differential expression than expected under neutrality (H-value ≥ 0.95, Figure 4b and Table S2), indicating that the observed pattern is due to local adaptation for these transcripts. Pop.LC had the highest number of DETs deviating from the neutral expectations (67% transcripts out of 2381), followed by Pop.G (~44% out of 3163), Pop.J (~49% out of 2679), and Pop.M (35% out of 2597).

Expression variation among individuals and populations

The statistical significance of difference between group means of expression values was assessed with a linear mixed model analysis for each transcript, evaluating the factors 'population' and 'Genotype'. For 414 transcripts, the means were statistically significantly different between populations but not between genotypes. The reverse was true for 10,201 transcripts. For 10,060 transcripts, the factors 'genotype' and 'population' explained the observed variation in gene expression. The remaining 12,228 transcripts had no significant p_{adj}-values for either of the factors.

Sequence based divergence

After applying the VariantFiltration criteria in the GATK SNP calling step, the resulting SNP set contained 414,546 variants distributed in 14,860 transcripts. These transcripts had an average of 28.2 SNPs per transcript. The vast majority (13,597 transcripts) was found to be biallelic and 1,083 transcripts were multiallelic (Table 1).

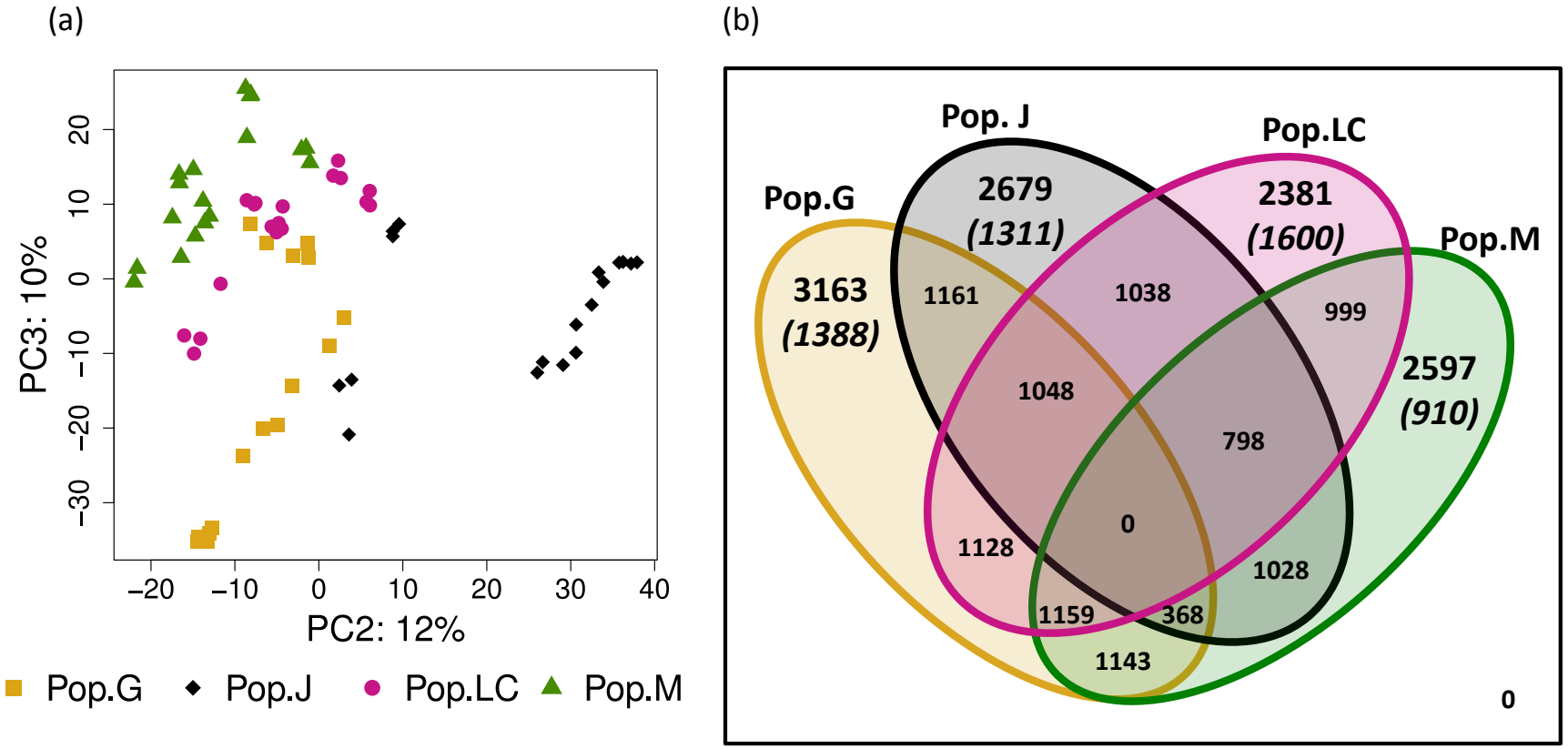


Figure 4: Gene expression patterns.

(a) Gene expression PCA of the four sampled populations: Pop.G (Lake Greifensee), Pop.J (Jordan Reservoir), Pop.LC (Lake Constance) and Pop.M (Müggelsee). Percentages on the X- and Y- axis indicate the percentage of variance explained by each principal component. (b) Venn diagram illustrating the number of differentially expressed transcripts (DET) between the four populations. Numbers in brackets indicate the number of transcripts deviating from the neutral expectations according to the DRIFTSEL analysis.

Table 1: Summary of SNP data.

"NonDET" refers to transcripts that were not significantly upregulated in any of the pairwise contrasts.

Population		all SNPs	biallelic SNPs	multiallelic SNPs
Pop.G DETs	Number of transcripts	1,369	1,259	110
	Number of SNP sites	34,525	34,320	205
	Average number of SNPs	25.21	27.25	1.86
Pop.J DETs	Number of transcripts	1,203	1,101	102
	Number of SNP sites	28,252	28,078	174
	Average number of SNPs	23.48	25.50	1.71
Pop.LC DETs	Number of transcripts	1,548	1,487	61
	Number of SNP sites	49,451	49,342	109
	Average number of SNPs	31.94	33.18	1.78
Pop.M DETs	Number of transcripts	1,087	992	95
	Number of SNP sites	36,772	36,598	174
	Average number of SNPs	33.82	36.89	1.83
NonDET	Number of transcripts	9,473	8,758	715
	Number of SNP sites	265,546	264,081	1,465
	Average number of SNPs	28.03	30.15	2.04
Total	Number of transcripts	14,680	13,597	1,083
	Number of SNP sites	414,546	412,419	2,127
	Average number of SNPs	28.23	30.33	1.96

A PCA was carried out based on a matrix of all biallelic SNP sites to illustrate the population structure among the four populations. Although PC1 explained the maximum variance (12%) (Figure 5a) and four distinct clusters corresponding to the populations were seen against PC2. PC2 and PC3 each explained 8% of the variance (Supplementary information Appendix S2). PC2 clearly separated the genotypes belonging to Pop.J from the remainder of the data (Figure 5a).

Inbreeding coefficient

The inbreeding coefficient values ranged from -1.19 to 0.22 for genotypes from Pop.G, from 0.08 to 0.17 for those from Pop.J, from 0.07 to 0.15 within Pop.LC, and from -0.06 to 0.32 for those belonging to Pop.M (Figure 5b). Within Pop.G, four out of the six genotypes were more heterozygous than expected; indicating outbreeding. In Pop.J, Pop.LC, and Pop.M (except genotype M9 therein), there were less heterozygous than expected, implying inbreeding.

Sequence evolution

To assess the respective contributions of random and non-random evolutionary events on DNA sequence divergence, we calculated the Tajima's D statistic for each transcript in the four populations. After phasing, we obtained 13,006 transcripts containing SNPs. Pop.LC had the highest number of transcripts (32.21%) with a negative D value ($D < 0$; $p \leq 0.05$) followed by Pop.G (30.45% transcripts), Pop.M (29.58% transcripts) and Pop.J (29.31% transcripts). Much fewer transcripts were found to have a significant positive Tajima's D value (Table 2): 1.26% transcripts in Pop.M, 1.20% transcripts in Pop.G, 1.13% transcripts in Pop.J and 0.66% transcripts in Pop.LC (Table S3).

The LOSITAN analysis identified 782 transcripts to be under diversifying selection, 1536 transcripts under balancing selection and 113 transcripts that were under balancing and/or diversifying selection (Table S4). LOSITAN results are described in detail in (Herrmann *et al*. 2017c).

The LOSITAN analysis identified 782 transcripts to be under diversifying selection, 1536 transcripts under balancing selection and 113 transcripts that were under balancing and/or diversifying selection (Table A1_T4). LOSITAN results are described in detail in (Herrmann *et al*. 2017c).

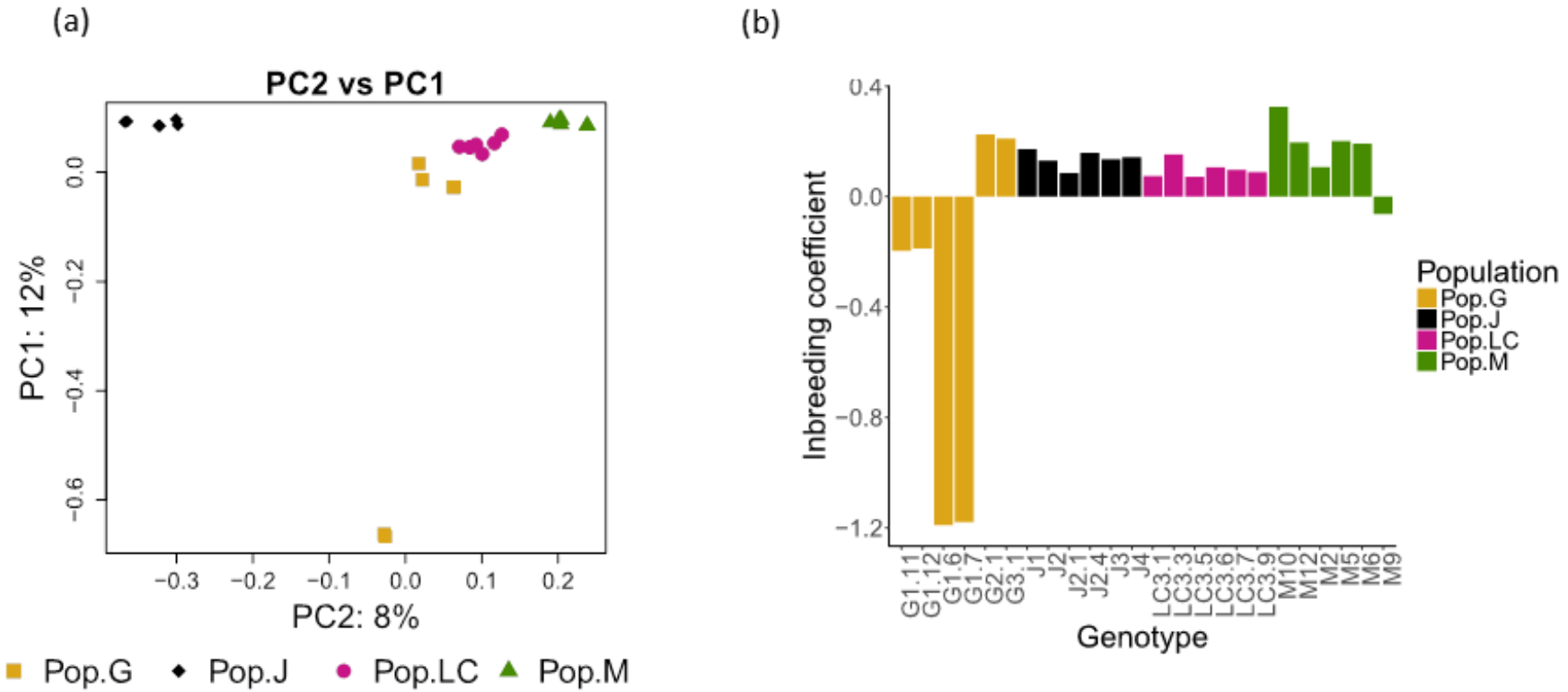


Figure 5: SNP patterns and heterozygosity.

(a) SNP PCA of the four sampled populations: Pop.G (Lake Greifensee); Pop.J (Jordan reservoir), Pop.LC (Lake Constance) and Pop.M (Müggelsee). Percentages on the X- and Y-axis indicate the percentage of variance explained by each principal component. (b) Barplot illustrating the inbreeding coefficient values for each genotype.

Table 2: Tajima's D test for selection.

D < 0: number of transcripts with a negative Tajima's D and thus likely under purifying selection; D > 0: number of transcripts with a negative Tajima's D and thus likely under balancing selection.

Population	D < 0	D > 0	Total
Pop.G	3,961	157	4,118
Pop.J	3,813	147	3,960
Pop.LC	4,192	87	4,279
Pop.M	3,848	164	4,012

Sequence vs. regulatory variation

The proportion of transcripts identified to be candidates for local adaptation at both sequence and regulatory level were visualized using a flow diagram (Figure 6). Among the 10,820 transcripts identified to be differentially expressed, ~46% showed signs of selection at the regulatory level according to DRIFTSEL. Of these, ~15% were identified as outliers under balancing and/or diversifying selection in LOSITAN. About 26% of these outliers had a significantly negative or positive Tajima's D value in at least one population, which might be attributed to selection but can also stem from other evolutionary processes such as population growth, reduction or subdivision, bottleneck events and migration.

Functional annotation

Of all transcripts, 66.5% had a BLAST hit to the nr database with an identity ≥ 50% and evalue ≤ 0.1; 91.4% transcripts of these BLAST hits shared homology with other *Daphnia* species. Among the DETs, 70.4% met this criterion (Supplementary information Appendix S3a, Table S5), and 92.3% of them were homologous to *Daphnia* sequences.

We were able to predict domains for ~50% of our transcripts. Among the DETs, a slightly higher proportion of transcripts, ~53%, had known protein domains (Supplementary information Appendix S3b, Table S5).

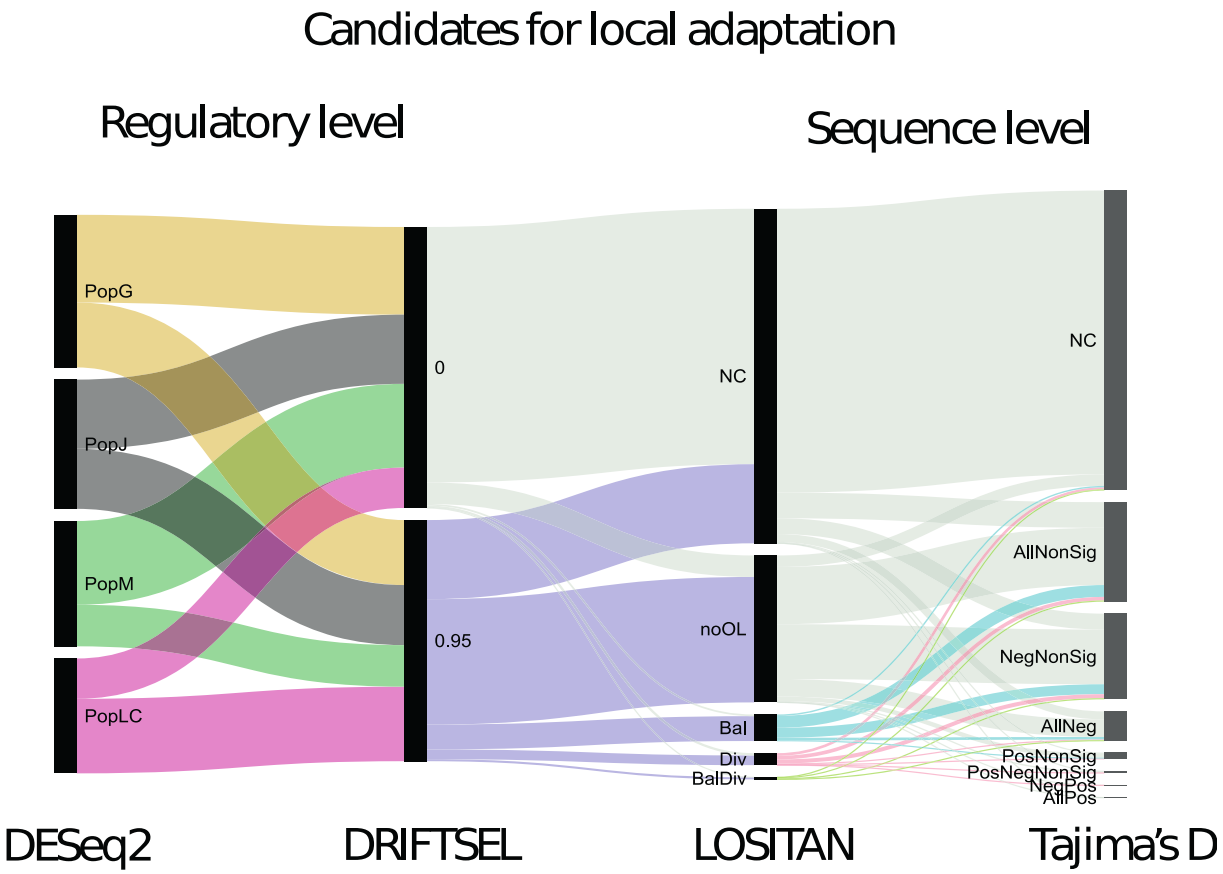


Figure 6: Flow diagram representing the proportion of transcripts that are candidates for local adaptation at the regulatory and sequence level.

Each analysis or "step" is represented by a vertical group of black rectangle bars, called nodes. The colored areas linking the nodes are called "flows". The **DESeq2** step contains four nodes: PopG (yellow), PopJ (black), PopLC (pink) and PopM (green), which represent the number of transcripts specifically upregulated in each of the four populations as identified by DESeq2 analysis. The **DRIFTSEL** step contains 2 nodes: 'H.value ≤ 0.95' (grey) and 'H.value ≥ 0.95' (purple). The **LOSITAN** step contains 5 nodes: 'NC' (grey) with transcripts without LOSITAN result (not calculated); 'noOL' (grey): transcripts where none of the SNPs in a transcript were identified as outliers; 'Bal' (cyan), transcripts containing at least one SNP that is under balancing selection; 'Div' (pink) transcripts containing at least one SNP under diversifying selection; and 'BalDiv' (pale green), transcripts containing SNPs that are under both balancing and diversifying selection. The **Tajima's D** step contains 8 nodes. Each node classifies the transcripts according to the obtained Tajima's D values. 'AllNeg' means that transcripts have a negative D value in all four populations; 'AllPos' means that transcripts have a positive D value in all four populations; 'AllNonSig' means transcripts have non-significant D values in all four populations; 'NegNonsig' means transcripts in the four populations have either a negative D value or a nonsignificant D value; 'PosNonsig' means transcripts in the four populations have either a positive D value or a nonsignificant D value; 'PosNeg' means transcripts in the four populations have either a positive or negative D value; 'PosNegNonsig' means transcripts in the four populations have either a positive or negative or an insignificant D value.

For identifying *Daphnia*-specific orthologs and those that share orthology with other arthropods, the orthoMCL data was classified into six categories (as described in the Methods section). 3,058 orthology clusters (of which 1,735 clusters contained DETs) were containing exclusively *D. galeata* transcripts, 985 clusters (of which 543 clusters contained DETs) contained only *D. galeata* transcripts and *D. pulex* genes, 651 clusters (including 224 DETs) contained only *D. galeata* and *D. magna* transcripts. 3336 orthoMCL clusters (of which 1239 clusters contained *D.galeata* DETs) contained all three *Daphnia* species used in the analysis. Furthermore, 12 clusters (4 clusters containing DETs) were containing *D. galeata* transcripts along with two other arthropods (*D. melanogaster* and *N. vitripennis*). In total, 4657 clusters (1586 clusters containing DETs) contained transcripts/genes for all five species (three *Daphnia* species and two insects) used in the present study (Supplementary information Appendix S3c, Table S5).

Assessment of assembly artefacts and inparalogs

In total, 3,325 DETs belonged to the "no-cluster DETs" category (Figure 7a), 5,574 DETs were exclusively occurring in orthoMCL clusters without DETs from different populations (1Pop). This vast majority was thus not further analyzed with regard to paralogy and assembly artefacts. The remaining 1,921 DETs were co-occurring with DETs from other populations in 716 orthoMCL clusters. Sequence divergence was calculated for every DET pair that co-occurred in a cluster. The divergence values ranged from 0.0 to 12.0 (Figure 7b). We cannot exclude that divergence values greater than 2 between sequence pairs arose from misassemblies. However, 16,752 sequence pairs (belonging to 671 clusters) had a divergence lower than our arbitrary threshold of 2, indicating that the transcripts were highly similar in their sequence and thus might constitute inparalogs or alternative transcripts for a gene. In this case, only genomic data would allow placing the transcripts and eventually assessing their status.

Gene Ontology enrichment analysis

GO enrichment analysis was performed on the candidate transcripts as identified from DRIFTSEL (Hvalue ≥ 0.95) and Tajima's D analyses. We observed an enrichment for several metabolic processes such as ATP binding, DNA binding, microtubule binding,

transporter activities and signaling pathways (Table S6) in both analyses in all population-specific sets. Specifically, in Pop.G, DRIFTSEL and Tajima's D analysis had five GO terms in common, in Pop.J, they had one GO term in common, in Pop.LC they had four GO terms in common and in Pop.M, they had seven GO terms in common.

Weighted Gene Co-expression network analysis

The WGCNA on 32,375 transcripts identified 29 co-expression modules (Figure 8) in the reference network (see Methods). We observed varying numbers of modules and transcripts clustered in each population-specific network (Table S7a-d). However, after assessing the conserved modules, where each population-specific network was compared to the reference network, 24 modules (out of 29) were well conserved (Zscore ≥ 10) among the populations. The conserved modules included 10,256 transcripts altogether, which is about 31% of all transcripts in *D. galeata,* with the largest module, 'turquoise' including 2,857 transcripts. Two modules (grey and gold) with uncharacterized and random transcripts contained 16,600 and 1000 transcripts, respectively. These results are consistent with the gene expression analysis which showed little differences between the populations.

Discussion

In this study, we describe an approach to distinguish between neutral and adaptive evolutionary processes at gene expression and DNA sequence level using *D. galeata* transcriptome data. We identified differentially expressed transcripts in each of the four populations. We also used the multivariate DRIFTSEL approach combining expression values and microsatellite data, to investigate the role of selection in shaping *D. galeata* differential expression profiles. Furthermore, we identified SNPs to understand the sequence level differentiation among the four populations. Finally, we annotated the functions of our candidate transcripts for local adaptation. This study is a first step towards description of polymorphisms in *D. galeata* possibly involved in phenotypic responses to environmental perturbations and as such promising candidates for future studies.

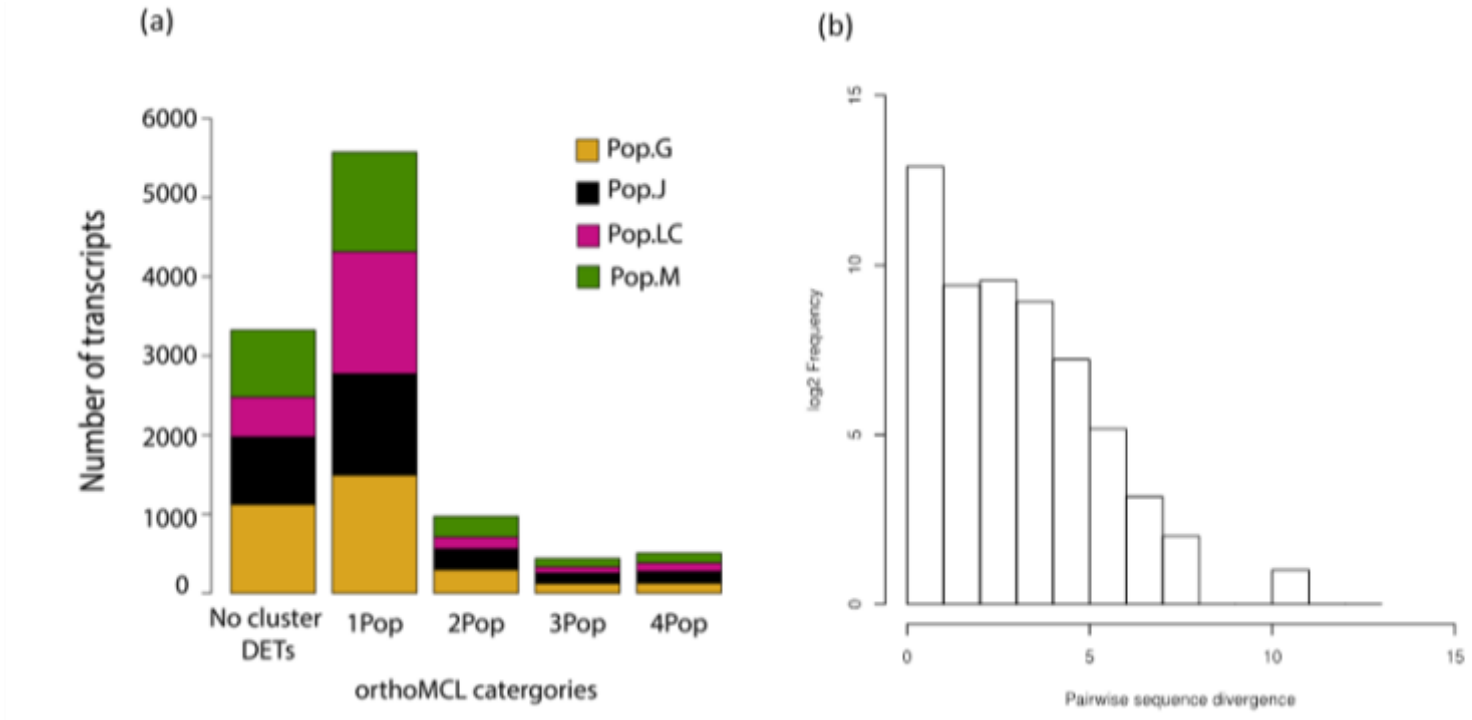


Figure 7: Assessment of assembly artefacts and inparalogs.

(a) Barplot showing the number of DETs co-occurring with DETs from other populations within an orthoMCL cluster. "No cluster DETs" refers to DETs not assigned to an orthoMCL cluster. 1Pop, 2Pop, 3Pop and 4Pop refer to DETs found in orthoMCL clusters containing at least one, two, three and four population(s) respectively. (b) Histogram of pairwise sequence divergence values calculated for all *D. galeata* sequences co-occurring in an orthoMCL cluster belonging to 2Pop, 3Pop and 4Pop categories.

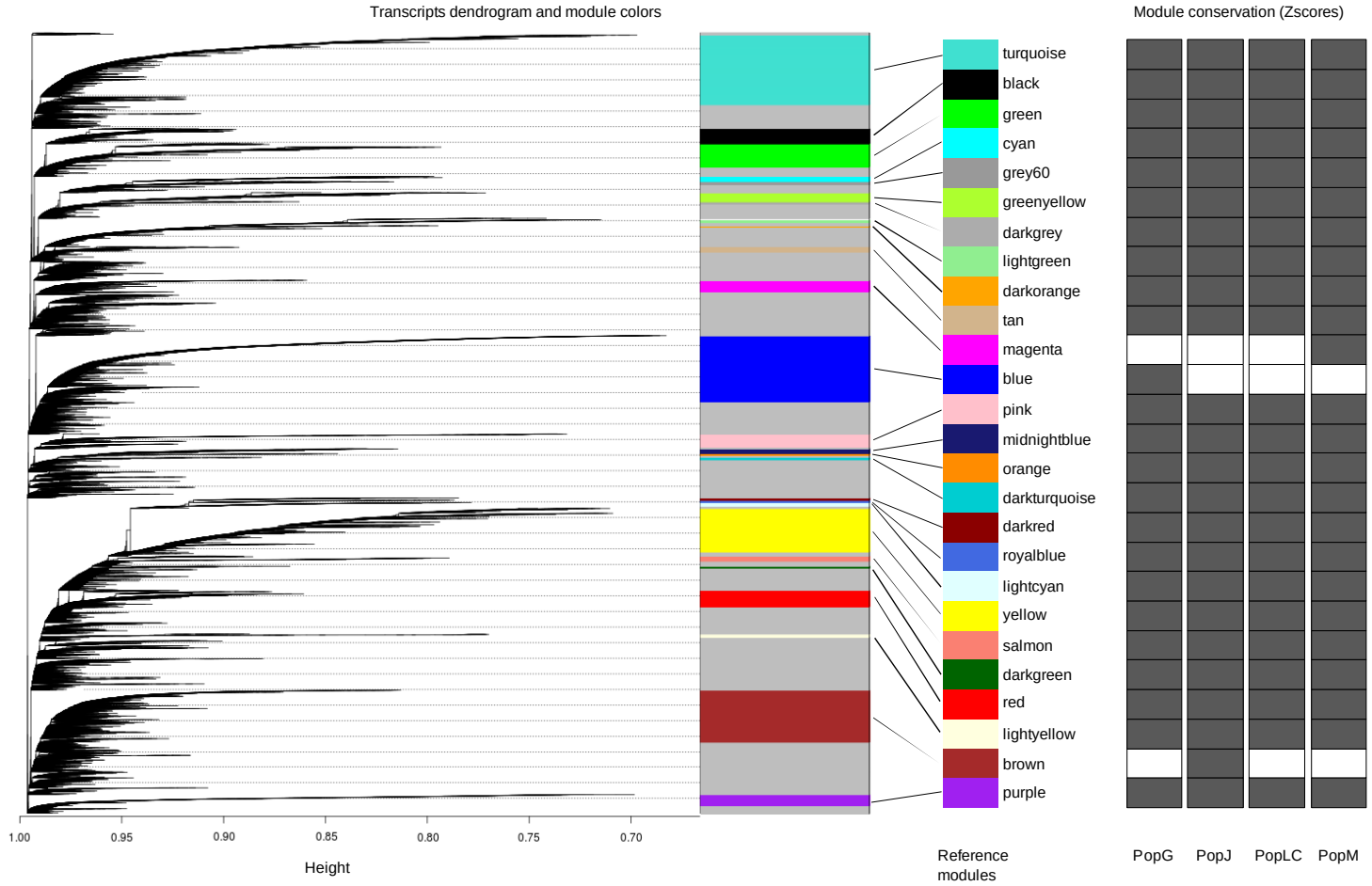


Figure 8:Cluster dendrogram of transcripts for the reference network in* Daphnia galeata*, with dissimilarity based on the topological overlap matrices (TOM).

The co-expression modules are colored in an arbitrary way in the WGCNA package, and the size of the bar is proportional to the number of transcripts in the module. The right hand side grid represents the module conservation in each population. Modules with a Z-score ≤ 10 are shown in white and modules with a Z-score ≥10 are colored in dark grey.

Population divergence at the sequence level

SNPs became the absolute marker of choice for molecular genetic analysis as the mining of polymorphisms is the cheapest source for genetic variability (Taillon Miller *et al.* 1998). Our PCA analysis on SNP data revealed four clear population clusters and our results are in agreement with a highly structured population model across the transcriptome. Although two of the genotypes (G1.6 and G1.7) from Pop.G were located outside the Pop.G cluster in the PCA plot, the populations were clearly distinguished and corresponded to the four lakes sampled. This pattern might be the result of several non-exclusive phenomena: initial founder effects, isolation-by-distance and genetic drift, and natural divergent selection, since the studied populations originate from lakes located in different ecoregions.

Genetic differentiation among populations of passively dispersed aquatic invertebrates is strong in most cases, despite the dispersal probability expedited by water birds and other vectors carrying their diapausing eggs (Mills *et al.* 2007; Munoz *et al.* 2016; Ventura *et al.* 2014). Population genetic differentiation has been observed even at small spatial scales (i.e., less than 1 km) in *Daphnia* (Hamrova *et al.* 2011; Yin *et al.* 2010). Additionally, the monopolization effect, a concept based on numerous previous studies on freshwater invertebrates (De Meester *et al.* 2002; Louette *et al.* 2007; Munoz *et al.* 2008; Ortells *et al.* 2013) might reinforce the population structure resulting from initial colonization event(s). Some evidence supporting this theory has been provided by Thielsch *et al.* (2015), who showed that novel genotypes are unlikely to successfully colonize a habitat if it already harbors an established population.

All the phenomena cited above have an impact on population structure across the genome, and might mask highly diverging loci resulting from natural selection. We assessed patterns of divergence at the sequence level through neutrality tests (Tajima's *D*). This suggested that all populations of *Daphnia* examined in this study had a substantial amount (~48% transcripts) of loci with an excess of low frequency polymorphisms (i.e., $D < 0$) relative to the neutral expectation. This pattern may result from positive selection, a bottleneck, or population expansion. It is consistent with previous observations in *Daphnia* from Lake Greifensee and Lake Constance (Brede *et al.* 2009) and crustacean zooplankton from Lake Constance (Straile 1998) which have all undergone historical bottleneck events.

Similarly, Lake Müggelsee, a large shallow lake, has undergone severe bottlenecks due to increased turbidity and because vegetation disappeared almost completely after the 1960s (Okun *et al.* 2005). One other explanation for the excess of rare alleles is selection against genotypes carrying deleterious alleles.

Although a high frequency of rare polymorphisms was observed in our analysis, there were few transcripts (~1.7% transcripts) that had a lower frequency of rare alleles (D > 0) in the four populations; indicating that some loci are either under balancing selection (where heterozygous genotypes are favored) or under diversifying selection (where genotypes carrying the less common alleles are favored). A lower frequency of rare alleles also occurs if there is a recent population admixture (Stajich & Hahn 2005). This argument is consistent with our inbreeding coefficient measures. Most of the genotypes in population G, as well as M9, were more heterozygous than expected. While genotype M9 from Müggelsee might be an exception, the pattern observed in Greifensee could be the consequence of outbreeding and/or high genetic variability in this population. This might also stem from past hybridization events (Brede *et al.* 2009). Although our primary checks for this with a handful of microsatellite markers did not flag the genotypes as having a hybrid origin, the use of high density markers such as the SNPs might have uncovered traces of introgression. Under the Hardy-Weinberg equilibrium, genotypes G2.1 and G3.1 from Pop.G, all genotypes in Pop.J and Pop.LC, and all genotypes except M9 in Pop.M show that the observed heterozygosity was less than the expected heterozygosity, which is an indication of lower genetic variability and inbreeding. Such low heterozygosity patterns at the individual level can be attributed to inbreeding (Keller 2002), but also due to a lack of variation in the source population, either caused by a small founder population size or a severe bottleneck during population history (Luikart *et al.* 1998b). Further, the ecology and growth dynamics of *Daphnia* populations might exacerbate the founder effects. After an initial hatching phase from the resting eggs bank and exponential population growth in the spring, clonal selection occurs throughout the growing season (Vanoverbeke & De Meester 1997b). Therefore, it is possible that only a few clonal lines contribute to the resting eggs population each year. However, while a reduced number of clonal lines might contribute to the yearly "archiving" of genetic diversity; two processes counteract the immediate diversity loss. First, the spring recruitment doesn't only rely on eggs from the

previous year but rather on a mixture (Vanoverbeke & De Meester 1997b), and might even integrate overwintering clones in larger permanent lakes (but see Yin *et al.* 2014 for an overview). Second, clonal erosion doesn't affect the same genotypes every year, leading to year-to-year heterogeneity, such as the one observed in the long term study by Griebel *et al.* (2016). Clonal erosion thus doesn't necessarily lead to a downward spiral of genetic diversity loss, and the high stochasticity of both clonal selection and hatching ensure a preservation of the genetic diversity in every habitat.

Gene expression variability and signals of selection

While the patterns observed at the sequence level tends to support the role of genetic drift, founder and monopolization effects in shaping the observed patterns, the results of our gene expression analysis delivered a mixed message. This was evident in the PCA based on the gene expression data, where no distinct clusters corresponding to populations are clearly visible. This observation was consistent with our network co-expression analysis which showed that the identified modules are conserved in all populations (Figure 8), with a few exceptions. The analysis of variance confirms this finding, with a relatively low number of transcripts for which the mean read counts differs significantly between populations and not between genotypes. While studies on differential expression in *Gliricidia sepium* (Chalmers et al. 1992) and *Arabidopisis halleri* (Macnair 2002) have observed substantial between population variance at the gene expression level, our results are consistent with several studies, for example, on fish (*Fundulus heteroclitus*; Whitehead & Crawford 2006a) and snails (*Melanoides tuberculata*; Facon *et al.* 2008) which showed large within-population variation. Additionally, numerous studies on life-history traits in *Daphnia* also report very high intrapopulation variability (Beckerman *et al.* 2010; Castro *et al.* 2007; Cousyn *et al.* 2001; Macháček 1991). A common garden experiment conducted on the very same clonal lines also showed a higher phenotypic variability within populations than among populations (Tams *et al.* 2018). Finally, the observed relative homogeneity in the gene expression profiles might be the consequence of high selective pressure on transcription regulation or canalization (Waddington 1942). Such canalization allows for storage of cryptic genetic variation that would be uncovered in stress response assessments. However, our experimental setup

was designed to avoid stress, and transcriptome characterization of the same genotypes under conditions mimicking predation, parasite or food stress, for example, might reveal a greater divergence between the populations.

Comparisons of the gene expression profiles for the four populations revealed a fair number (~8%) of *D. galeata* transcripts to be significantly exclusively upregulated in one given population compared to all others. Although all populations showed similar numbers of differentially upregulated transcripts, when considering those which are probably under directional selection, the picture changed. After applying the DRIFTSEL approach, Pop.LC had the highest number of transcripts directionally selected based on their expression levels and Pop.M had the lowest number. Pop.G and Pop.J had nearly similar numbers of transcripts under directional selection. A study on adaptive differentiation in seagrass (Jueterbock *et al.* 2016) that compared Northern and Southern seagrass samples under thermal stress showed that natural selection was the most straightforward explanation for nearly 1% of all differentially expressed genes. For other genes that were differentially expressed in the seagrass study, parallel adaptation to different habitats was observed along both the American and European thermal clines. However, more analysis (such as McDonalds Kreitman test) and a combination of other factors (such as phenotype differences among populations) is required to make such inferences for parallel adaptation in *D. galeata* populations.

Sequence vs. regulatory variation in Daphnia galeata

Correlating expression profiles with sequence divergence helps to identify transcripts that are potentially under the influence of local adaptation at both gene expression and sequence level. Linking gene expression profiles with sequence polymorphisms and their associated functions aids in understanding the genetic basis of adaptation as seen in the desert adapted mouse (*Peromyscus eremicus*; MacManes & Eisen 2014) and in the Patagonian olive mouse (*Abrothrix olivacea*; Giorello *et al.* 2018). Our results revealed ~30% of the transcripts to share divergence at both sequence and regulatory level (Figure 6). There are two possible explanations for the observed differences in sequence and regulatory level variation (Hodgins *et al.* 2016). The first is that there is an increase in the rate of fixation due to transcripts under positive selection and divergence in expression

patterns. For example, variation in gene expression might lead to selection for sequence variation to improve the functional role of the transcript in its altered role (Hodgins *et al.* 2016). A second explanation is that the differentially expressed transcripts may experience reduced negative selection in one or all four populations. For instance, higher transcript expression is associated with greater negative selection. Hence a reduction in transcript expression in one population compared to others may be accompanied by relaxation of selection in that population.

GO enrichment analysis on the candidates identified at the sequence (Tajima's D) and expression (DRIFTSEL) level were enriched for metabolic and cellular processes. These findings suggest that there may be a hierarchical activation of general mechanisms of stress responses at the metabolic and cellular level. This observation is concordant to another study (Orsini *et al.* 2017) on *D. magna*. In this study, *D. magna* were subjected to several environmental perturbations and the GO enrichment analysis revealed a general stress response rather than ontologies specific to local adaptation. Since the present study is without any laboratory induced stressor, further studies in *Daphnia* subjected to one or multiple environmental stressors would be helpful in pinpointing stress specific responses. Further, no GO term annotation was available for ~31% of the transcripts, and we cannot therefore reach conclusive results. This highlights the need for new and complementary resources for *Daphnia* genomics research, and a general improvement of the existing annotation.

Gene annotation and evaluation of inparalogs

Gene annotation is quite challenging in organisms lacking reference genomes, and functional annotation then relies on the availability of transcriptomic sequences from the closest available taxon. In this study we were able to annotate 66.5% of the transcripts using BLAST analysis (Supplementary information Appendix S3a). However, many of the transcripts were homologous to a *D. pulex* "hypothetical protein", likely because (i) they are similar in function to non-coding regions or pseudogenes or (ii) novel coding transcripts that are yet to be functionally characterized (Vatanparast *et al.* 2016). Furthermore, we were able to predict domains for 80% of the transcripts using Pfam analysis (Supplementary information Appendix S3b, Supplementary Table S4). Our

orthoMCL results (Supplementary information Appendix S3c, Supplementary Table S4) showed that several (~45%) of the *D. galeata* transcripts were orthologous to one or all species of *Daphnia* used for comparison, indicating that the genes/transcripts have all evolved from a common *Daphnia*-specific ancestral gene via speciation. In addition to this, ~25% of *Daphnia* genes/transcripts are orthologous to two insect species (*D. melanogaster* and *N. vitripennis*). Our level of unannotated transcripts is similar to results reported from other organisms lacking extensive genomic resources, for example, from plants like field pea (*Pisum sativum;* Sudheesh *et al.* 2015), chick pea (*Cicer arietinum;* Kudapa *et al.* 2014), and winged bean (*Psophocarpus tetragonolobus*; Vatanparast *et al.* 2016). This limited our interpretation of the functional role of *Daphnia* transcripts and thereby their associations to known ecological stressors. A second issue raised when lacking a reference genome is that it might be difficult to tease apart inparalogs created by duplication events, isoforms and even misassemblies; leading to an artificially inflated number of similar sequences for each distinct gene in the transcript set. Only ~18% of the population specific DETs had one or more putative paralogs also identified as differentially expressed in at least one other population. For DETs from two or more populations that co-occurred in orthoMCL clusters, we were able to distinguish between actual paralogs (transcript pairs that had a sequence divergence value > 2, Figure 5b) and transcripts with sequence divergence value < 2.Genomic information is now required for this species in order to accurately assign transcripts to genes and correctly assess whether two different populations might indeed express different gene copies with similar functions

Future directions and conclusions

In summary, we described here an approach that combines both transcriptomic expression profiles and sequence information to understand local adaptation in *D. galeata*. Although the set of transcripts contributing to population divergence at the sequence and the expression level differ, both levels constitute alternative routes for responding to selection pressures (Pai *et al.* 2015); showing that these transcripts can contribute to local adaptation and paving way for future research. From our functional analysis, it was evident that most of our transcripts were *Daphnia* specific although they had hypothetical functions. To understand the function of the hypothetical transcripts in

D. galeata and their response to environmental perturbations, a comparative approach using the gene expression data from numerous other *Daphnia* studies should be used. Although we noticed correlations between expression patterns and sequence divergence for the *D. galeata* transcripts, we lack genomic and phylogenetic information. This information may help "bridge the gap" for understanding the relative roles of positive or negative selection in driving coding sequence and gene expression divergence.

Data accessibility

The raw sequence reads used for this study as well as the experimental set up for the analysis of differentially expressed genes are available on ArrayExpress (https://www.ebi.ac.uk/arrayexpress; Accession no.: E-MTAB-6144).

The raw read counts used as input for differential transcript expression , results for the pairwise contrast analysis conducted in DESEq2 ,and the number of variants per sample before and after filtering, number of variant sites per transcript are all available on DRYAD in Tables S10, S11 and S3-S4, respectively (https://datadryad.org//resource/doi:10.5061/dryad.p85m5). The VCF file will be made available on Dryad Digital Repository and accession numbers will be updated.

Supporting Information

The following figures are found in the "Appendix for Chapter 1" section.

Figure A1_F1: Gene expression PCA for the first three principal components

Figure A1_F2: SNP PCA for the first three principal components

Figure A1_F3a-c: Pie charts showing functional annotation using BLAST, Pfam and orthoMCL analysis.

The following tables are found in the enclosed CD-ROM in the folder "Appendix_Chapter1" or from the online version of the article upon publication.

Table A1_T1: Library preparation and sequencing information along with principal component coordinates for the first three axes as obtained from gene expression analysis.

Table A1_T2: DRIFTSEL values for the differentially expressed transcripts.

Table A1_T3: Population-wise Tajima's D values.

Table A1_T4: LOSITAN outlier test values to identify loci under selection.

Table A1_T5: Functional annotation for candidate transcripts of local adaptation.

Table A1_T6a-c: Population specific GO enrichment terms using DRIFTSEL and Tajima's D analysis.

Table A1_T7a-d: Number of transcripts clustered in each module as detected by WGCNA for PopG, PopJ, PopLC and PopM, respectively.

Acknowledgements

We would like to thank A. Jueterbock and O. Ovaskainen for suggestions and help on the DRIFTSEL implementation, and four anonymous reviewers as well as Luisa Orsini for their comments on earlier versions of the manuscript. This study was financially supported by an individual grant from the Volkswagen Stiftung (to MC). This work benefits from and contributes to the *Daphnia* Genomics Consortium.

Chapter 2: *Daphnia*Stressdb: Taking advantage of a decade of *Daphnia* '-omics' data for gene annotation

Suda Parimala Ravindran, Lisa Göttslisch, Jennifer Lüneburg, Verena Tams and Mathilde Cordellier

Abstract

Gene expression patterns help to characterize and measure the effect of environmental perturbations at the cellular and organism-level. Complicating interpretation is the presence of uncharacterized or "hypothetical" gene functions for a large percentage of genomes. This is particularly evident in *Daphnia* genomes, which contains many regions coding for "hypothetical proteins" and are significantly divergent from many of the available arthropod model species but might be ecologically important. In this meta-analysis, we developed a gene expression database, the *Daphnia* stressor database (http://www.daphnia-stressordb.uni-hamburg.de/dsdbstart.php), built from 90 published studies on *Daphnia* gene expression. Using a comparative genomics approach, we used the database to annotate *D. galeata* transcripts. The extensive body of literature available for *Daphnia* species allowed to associate stressors with gene expression patterns. We were able to identify at least one stressor for about 4684 *D. galeata* transcripts. We believe that our stressor-based annotation strategy allows for better understanding and interpretation of the functional role of the understudied hypothetical or uncharacterized *Daphnia* genes, thereby increasing our understanding of *Daphnia*'s genetic and phenotypic variability.

Keywords: *Daphnia* – gene expression – stress -comparative genomics – functional annotation

This version of the chapter has been submitted to *Nature Scientific Reports*.

Introduction

Environmental health is an important determinant of human health (Myers *et al.* 2013) and natural habitats are under increasing threat from human activity. Industry and private households release chemicals in the environment, and human activity causes habitat fragmentation and warming, thus affecting ecosystems in their entirety. A wide diversity of organisms undergo changes in behaviour, morphology and physiology by responding to biotic and abiotic factors in their environments (DeWitt & Scheiner 2004; Harvell 1990; Karban 1989; Sultan 2000).

When an organism is subjected to one or multiple environmental perturbations, its gene expression and metabolic profiles are altered. Gene expression patterns provide clues about the biochemical pathways that are affected by a specific stressor. Organisms under stress might show unique signatures of gene expression patterns associated with a stress response as seen for example in plants and rat liver (McMillian *et al.* 2004). These unique patterns of gene expression can be used as indicators of specific environmental stressors.

Quantifying and comparing the gene expression profiles before and after exposure to a stressor is the most widely used approach to understand the genetic consequences of stressors. The central goal of most gen-/transcript-omic experiments is to obtain sets of differentially expressed genes and interpret the observed patterns. Complicating interpretation is uncharacterized gene function for a large percentage of genomes. This is particularly evident in crustaceans which are biologically diverse organisms and have been the subjects of investigation for hundreds of years because of their important roles in ecology, evolution and ecotoxicogenomics (Stillman *et al.* 2008). Amongst crustaceans, gene prediction tools in *Daphnia* genomes identified many coding regions but they are substantially divergent from many of the available arthropod model species, and annotation beyond "hypothetical protein" was not possible. Although *Daphnia*'s ecology is intensively studied, little is known on the molecular basis of responses to environmental stressors.

Daphnia is an ecologically important model organism distributed throughout the world in a variety of habitats ranging from lakes to ponds and from hypersaline to freshwater systems. They are primary grazers of algae, bacteria and protozoans and primary forage for fish. As a consequence, *Daphnia* are considered to be the sentinel

species of freshwater environments where their decline is proportional to the presence of environmental perturbations (Dodson & Hanazato 1995). Several studies (Garbutt & Little 2014; Gorr *et al*. 2004; Lampert *et al*. 2012; Weiss *et al*. 2015) in *Daphnia* exist which quantify their gene expression profiles in response to a stressor which affect both phenotype and metabolism. These observations have become more feasible with the advent of *Daphnia* genomics, which have seen rapid advancements in the last decade with the availability of the first genomes of *D.pulex* (Colbourne *et al*. 2011), *D.magna, D.similoides* and transcriptome of *D. galeata* (Huylmans *et al*. 2016). The wFLEABASE (Colbourne *et al*. 2005) developed by the *Daphnia* Genomics Consortium (DGC) currently serves as a hub for obtaining gen-/transcript-omic *Daphnia* data. However, a *Daphnia* repository dealing with experimentally validated gene expression and its response to environmental stressors is lacking.

By taking advantage of comparative genomics approaches, in the present meta-analysis we wanted to identify transcript-specific stressors in *D. galeata* using the functional annotation data available from a previous study (Huylmans *et al*. 2016). We observed that most of the *D. galeata* transcripts annotated functionally were "hypothetical" or the function was "unknown". To overcome this challenge, we identified studies dealing with *Daphnia* gene expression using three different search strategies. We manually mined the data to identify differentially expressed genes and transcripts for each stressor. Using a homology approach, we were able to annotate *D. galeata* transcripts to *Daphnia* specific genes in literature known to respond to one or more stressors through regulatory changes. We further identified transcripts/genes that were responding to many stressors thereby forming a general stress response and also identified stress-specific transcript/genes.

The results of our meta-analysis were implemented into a database called *Daphnia* stressor database (http://www.daphnia-stressordb.uni-hamburg.de/dsdbstart.php). To the best of our knowledge, this meta-analysis is a first attempt to integrate the wealth of information readily available by using a comparative genomics approach. We believe this long overdue resource will be of use to researchers working in the fields of ecotoxicology, environmental health, evolutionary biology and environmental genomics, and serve as a basis for further investigations in *Daphnia* research.

Methods

Identification and selection of eligible **Daphnia***-specific gene expression datasets for meta-analysis*

We used three different strategies to retrieve literature that analyzed gene expression data in *Daphnia*.

a. *Literature search based on functional annotation*

   Using the functional annotation data obtained from a previous study (Huylmans *et al.* 2016) on *D. galeata*, we searched for literature using the keywords "*Daphnia*" + "stress" along with the predicted function for each transcript.

b. *Literature search based on keywords in Mendeley*

   We used keywords "*Daphnia*" + "Gene Expression" in Mendeley reference manager and obtained sets of *Daphnia* specific gene expression papers.

c. *Twitter based search*

   We automatically retrieved all literature posted in @wtrflea_papers twitter handle using a python script (available upon request), and subsequently retained only studies about gene expression.

A consolidated literature list was created with studies obtained from all three approaches mentioned above. Studies were excluded from the analysis for the following reasons: different language than English, no traceable gene names, or only primer sequences instead of gene IDs. The following information was extracted from each identified study in the consolidated list: (i) gene IDs that are differentially expressed in response to a stressor or condition; (ii) method used for the analysis; (iii) stressor(s) used in the study; (iv) study species.

Sequence retrieval

Each retrieved gene or protein ID identifier was manually checked to see if they followed the NCBI (eg: AB021136) or Uniprot (eg: E9FXA0) or wFLEABASE (eg: DAPPUDRAFT_224348) nomenclature. Sequences for all obtained gene IDs were then retrieved from respective databases (NCBI / Uniprot / wFLEABASE) using the trial version of iMacros tool (https://imacros.net/overview/), which is an automated tool for web scraping and data extraction. For gene IDs that followed a *D. magna* ID nomenclature (e.g.:

Dapma7bEVm000787t1), sequences were directly obtained from the CDS file available on wFLEABASE (http://server7.wfleabase.org/genome/Daphnia_magna/openaccess/genes/Genes/earlyaccess/dmagset7finloc9b.puban.cds.gz).

For microarray-based studies on *D. magna*, the array accession number (eg: A-GEOD-16518) cited in the publication was queried in Gene Expression Omnibus (Edgar *et al.* 2002) (GEO). Reporter sequences were retrieved and a local BLAST was performed against the *D. magna* CDS sequences mentioned above. Hits that had a similarity ≥ 90% were considered for the next steps in the analysis.

All retrieved sequences were manually sorted into a "protein" or "nucleotide" category using an in-house perl script (available upon request).

Local sequence database creation and BLAST analysis

Two fasta files, one with all nucleotide sequences and one with all protein sequences were created. Redundant sequences from the two fasta files were excluded using the sequence cleaner biopython script (https://biopython.org/wiki/Sequence_Cleaner) with the default parameters. Thus, two non-redundant local databases were created with 28259 nucleotide sequences and 102 protein sequences for BLAST (Altschul *et al.* 1990) analysis.

Each of the 32903 *D. galeata* transcripts was used to query the local non-redundant nucleotide (tblastx) and protein (blastx) database created with the gene IDs obtained from literature. Hits with evalue (eval) ≤ 0 and identity percentage ≥ 50% were considered and the stressors corresponding to their respective subject ID was assigned to the *D. galeata* transcripts.

A general workflow of the entire meta-analyses is represented in Figure 9.

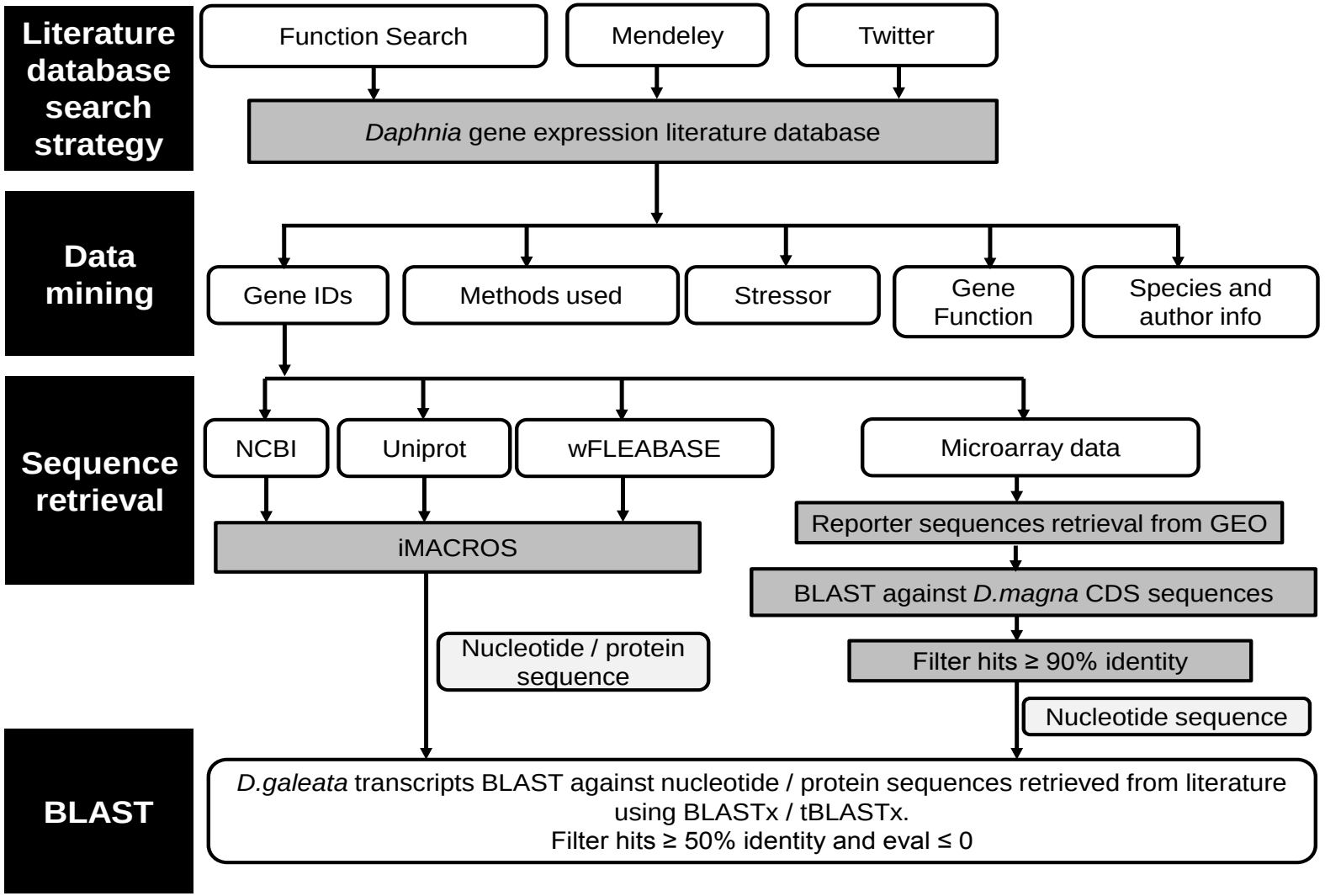


Figure 9: Workflow of Daphnia-specific gene expression meta-analysis.

Database web interface

Daphnia stressor database (http://www.daphnia-stressordb.uni-hamburg.de/dsdbstart.php) is built on Mariadb (10.0.31-MariaDB) at the backend whereas the frontend is built using HTML, PHP and CSS3. To allow users identify gene-specific stressors in *Daphnia* using a homology approach, we implemented a BLAST function on the database using SequenceServer (Priyam *et al*. 2015).

Statistical analysis

All statistical analyses were performed using custom scripts (available upon request) in R (R Development Core Team 2008).

Results and Discussion

Number of studies obtained from each search strategy

A total of 676 studies were obtained from all the three search strategies as described (in methods section). Of them, 47 studies came from FunctionSearch method, 74 studies came from Mendeley search and 555 came from automatic retrieval using twitter. We excluded studies not dealing with gene expression in *Daphnia* and studies from which information was not retrievable (due to language restrictions, intellectual property rights, primer sequences and/or gene names instead of IDs were mentioned). After screening the literature database with the above-mentioned criteria and checking for duplicates, gene expression data specific to *Daphnia* was retrievable from 90 studies (Table A2_T1) and used for the present meta-analysis (Figure 10a).

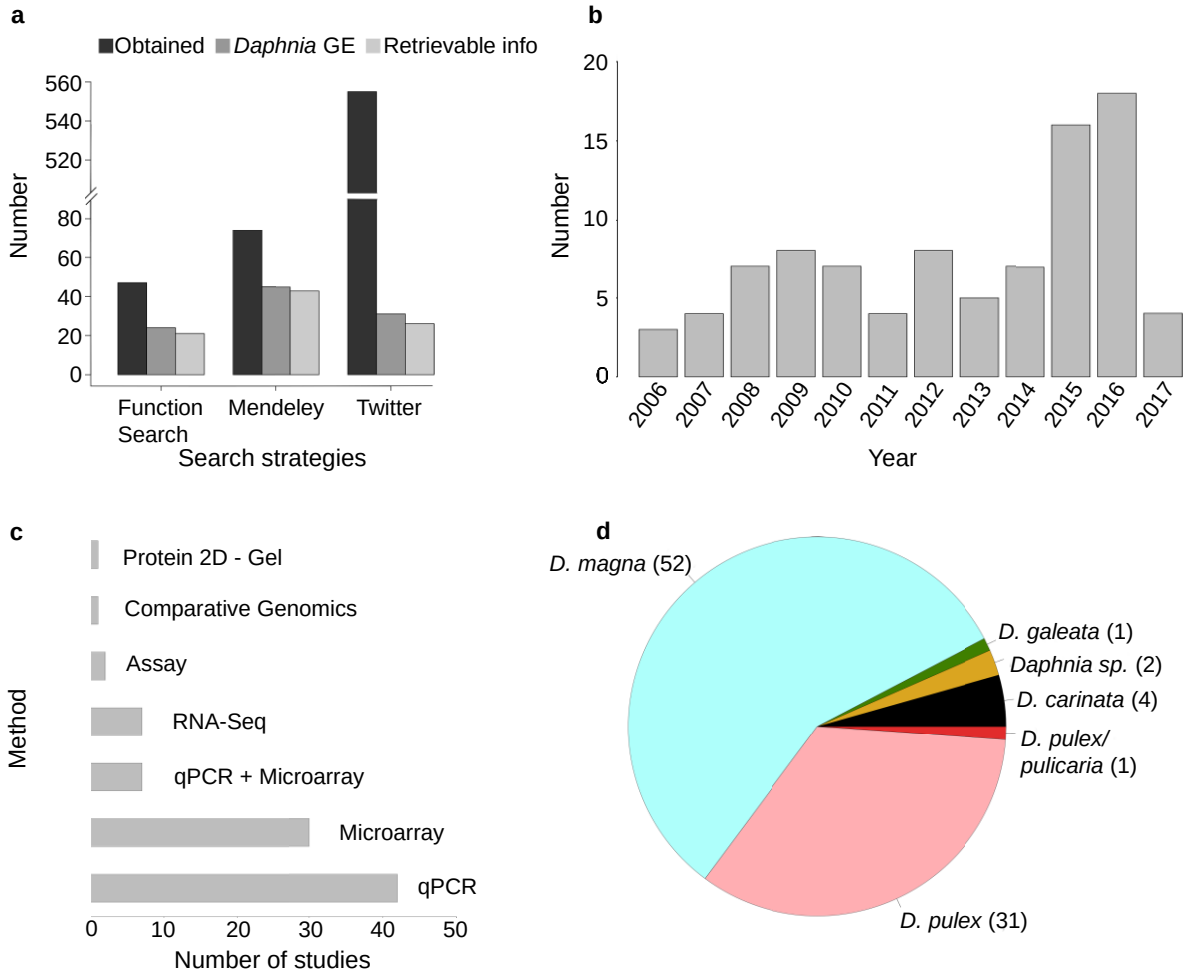


Figure 10: Literature statistics for meta-analysis.

(a) Barplot indicating the number of studies obtained for each search strategy used in this meta-analysis. Bars in black represent all literature obtained in each search strategy. Bars in dark grey represent the number of studies specific to *Daphnia* and gene expression; bars in light grey represent the number of studies from which data was retrievable for analysis.
(b) Timeline for studies on gene expression in *Daphnia*.
(c) Barplot showing the number of studies available for different analysis method used in the literature.
(d) Pie chart showing the number of studies available for each *Daphnia* species.

Timeline of literature, species information and methods used

After limiting the literature database to gene expression studies in *Daphnia*, we obtained literature from the year 2006 until mid-2017 (Figure 10b). About 43 studies used qPCR, thus limiting their analysis to only a few genes (Figure 10c). After the advent and further development of DNA microarray at the dawn of the millennium, researchers took advantage of the technology to perform large-scale gene expression studies. In our survey, 30 studies used microarray technology to analyze gene expression patterns in *Daphnia*. Seven studies combined both microarray and qPCR approaches. With the advent of high-throughput sequencing (HTS) techniques and affordable sequencing costs, *Daphnia* researchers used RNA-seq to sequence whole genomes and transcriptomes, thereby making the *Daphnia* genome a reality. In this meta-analysis, we found 7 studies that dealt with HTS data. Very few of the studies (3 studies) in the literature database used comparative genomics approaches, biochemical assays and 2D-gel electrophoresis techniques. The majority of the studies (52 out of 90; Figure 10d) dealt with *D. magna*, followed by *D. pulex* (31 out of 90) and very few on other *Daphnia* species like *D. carinata* (4 studies), *Daphnia sp.* (reviews, 2 studies), *D. galeata* and *D. pulicaria* (one study each).

Daphnia ***genes associated to a stressor in literature***

After retrieving the gene IDs regulated in response to a stressor from literature, ~21% of the 30939 genes (Colbourne *et al.* 2011) with a DAPPUDRAFT id (e.g.: DAPPUDRAFT_xxxxxx) present in *D. pulex* were associated to a stressor (Figure 11a). Among them, ~84% of the genes were affected by one stressor, 13% of the genes by two stressors, 2% by three stressors and less than 0.5% between 3 and 9 stressors.

For *D. magna*, ~25% of the 29127 genes were associated to a stressor that came from microarray studies (Figure 11b). Among them, ~51% of the *D. magna* genes associated to a stressor were responding to one stressor, ~23% to two stressors, ~9% to three stressors and less than 5% of genes were responding between 4 and 31 stressors.

To represent the shared response between stressors, we used a circularized plotting using circlize package (Gu *et al.* 2014) in R. Every stressor field represents the list of genes shown to be differentially expressed. A line linking one stressor field to another implies that a gene is differently expressed for both stressors. The link density thus makes the

non-specificity of some regulatory responses. For *D. magna*, stressor fields were often highly linked to other stressors (Figure 12). This leads to the conclusion that the genes might be showing a general stress response. In *D. pulex* (Figure 13), some stressors showed very few links (i.e. temperature, light: dark cycle, phosphorous). In this case, the stressor response might be specific and even used for environmental diagnosis.

A few studies did not use the "DAPPUDRAFT" or "Dapmaxxxxxxxxxx" nomenclature, yet they were dealing with stressors associated to *D. pulex* and *D. magna,* respectively. These genes and their associated stressors were visualized using bubbles (Figures 11, 12), where each bubble corresponds to one stressor and their size is proportional to the number of genes associated to that stressor.

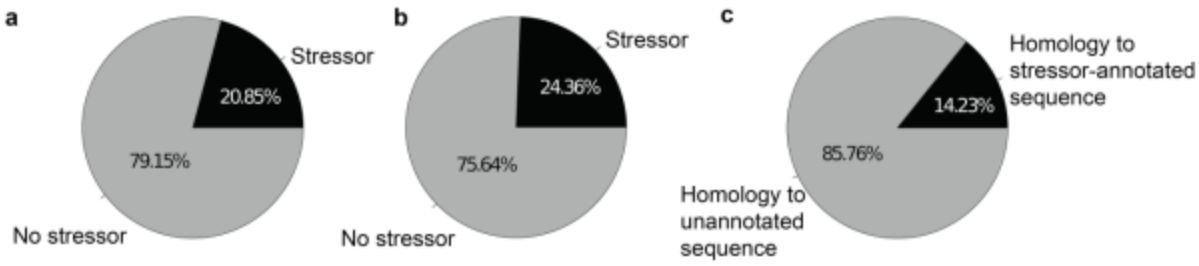


Figure 11a-c: Pie charts showing the percentage of genes associated to a stressor in the literature database.

(a) *D. pulex* (with DAPPUDRAFT nomenclature IDs); (b) *D. magna* (with Dapmaxxxxxxxx nomenclature IDs); (c) *D. galeata* (test case used in the present analysis).

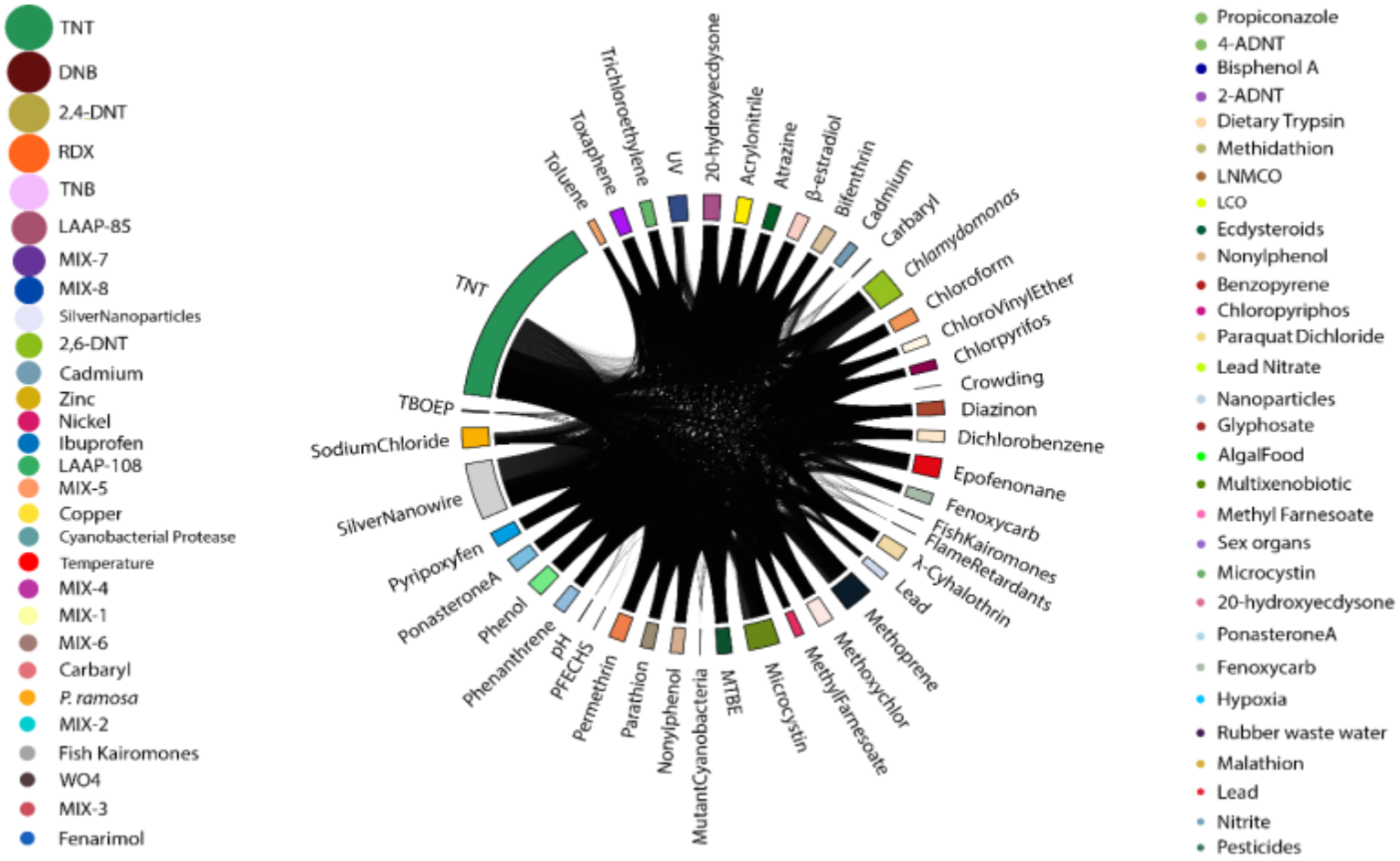


***Figure 12: Circular and bubble plot showing the number of genes associated to one or more stressors in* D. magna.**

Every stressor field on the circular plot represents the list of genes identified to be differentially expressed, and its size is proportional to the number of genes. A line linking a field to another one means that a gene is differently expressed for the other stressor as well. For genes that followed other gene ID nomenclature for *D. magna*, stressors are represented through bubbles on either side of the circular plot, and the bubble size is proportional to the number of genes.

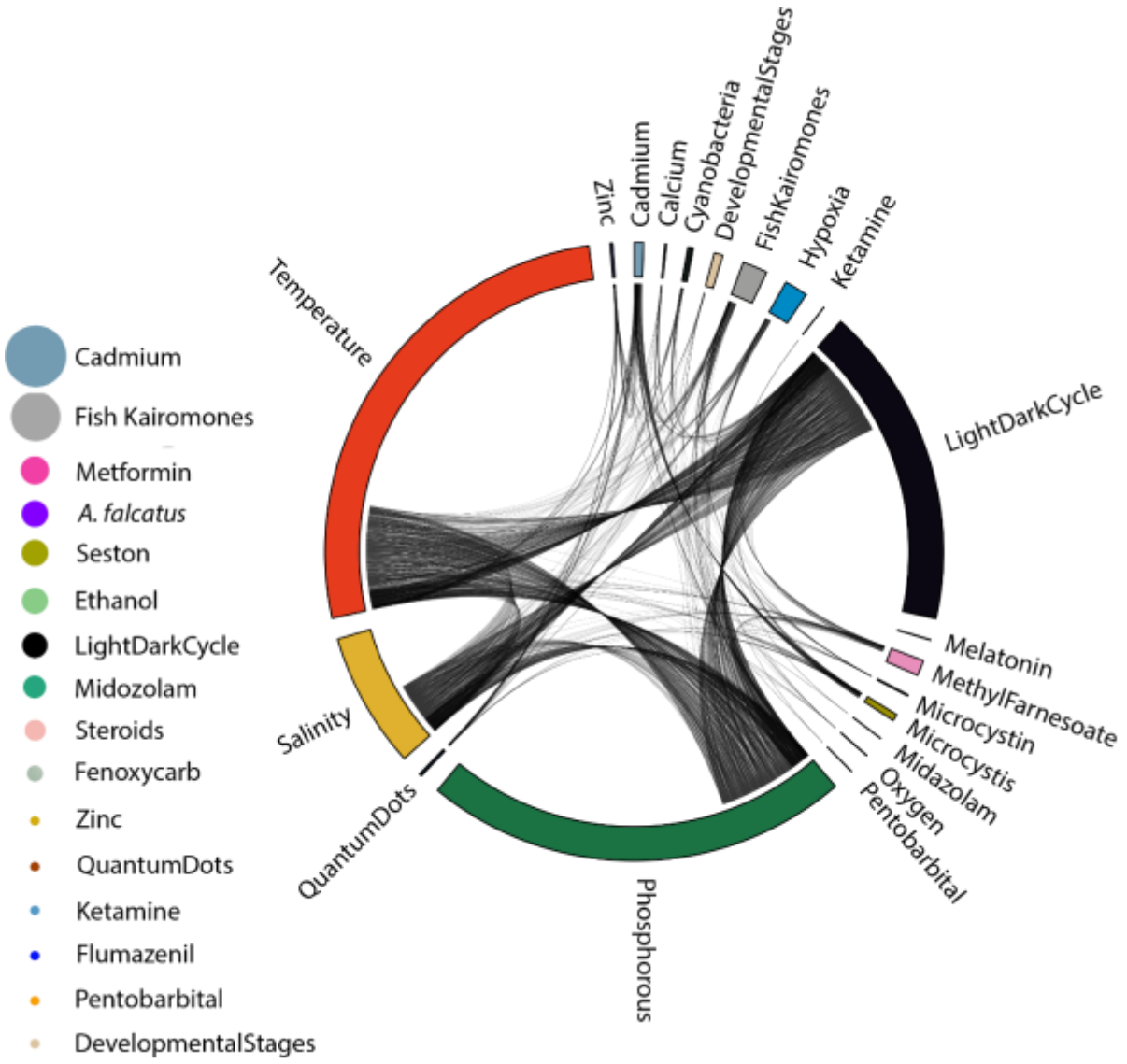


***Figure 13: Circular and bubble plot showing the number of genes associated to one or more stressors in* D. pulex.**

Every stressor field on the circular plot represents the list of genes identified to be differentially expressed, and its size is proportional to the number of genes. A line linking this field to another one means that a gene is differently expressed for this stressor as well. For genes that followed other gene ID nomenclature for *D. pulex*, stressors are represented through bubbles on the left side of the circular plot, and the bubble size is proportional to the number of genes.

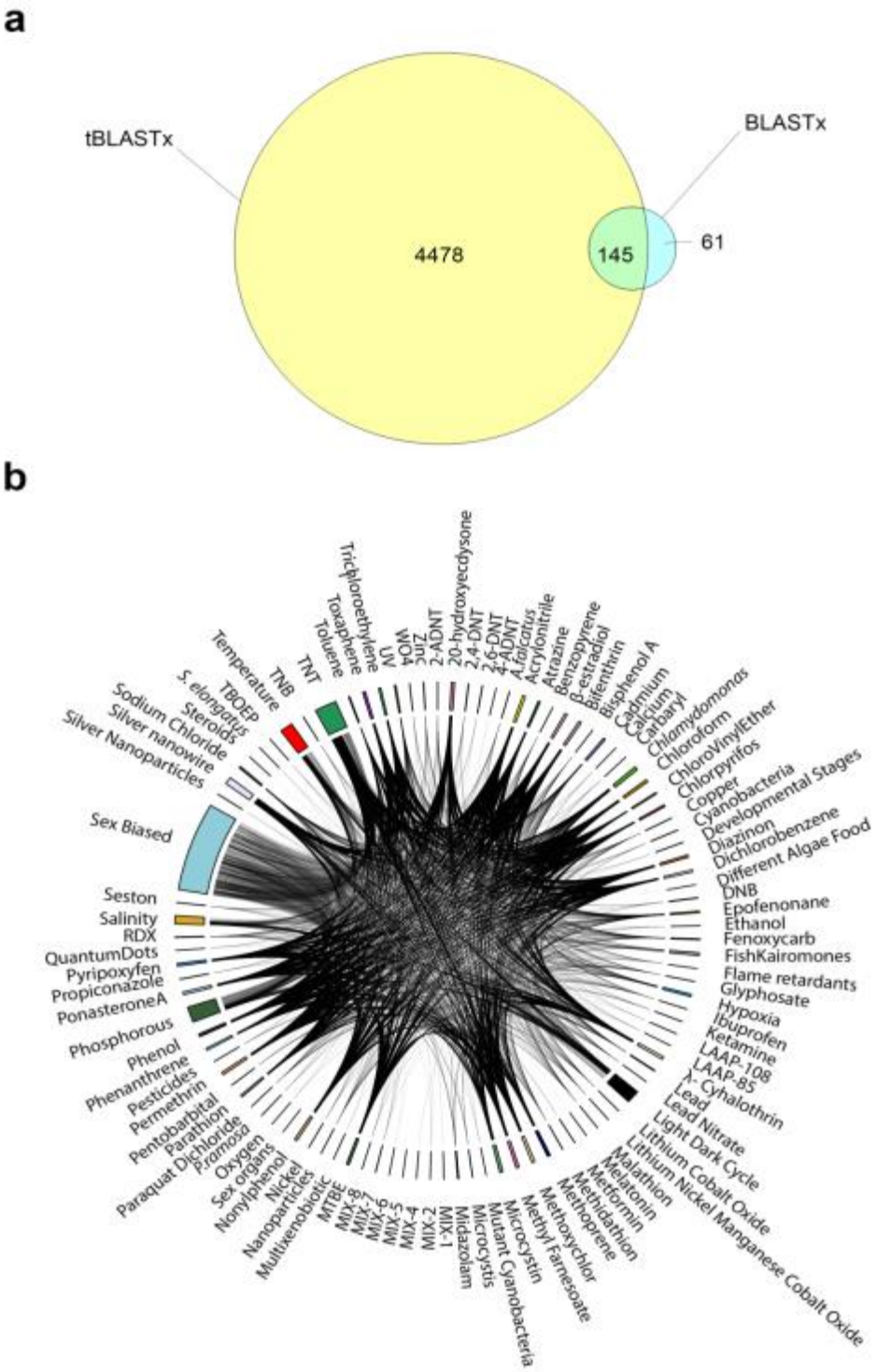


Figure 14: D. galeata transcripts associated to a stressor using homology approach.

(a) Venn diagram representing the overlap between blastx and tblastx analysis used for annotating *D. galeata* transcripts. (b) Circular plot showing the number of genes associated to one or more stressors in *D. galeata*. Every stressor field represents the list of genes identified to be differentially expressed. A line linking this field to another one means that a gene is differently expressed for this stressor as well.

D. galeata ***genes associated to a stressor using homology approach***

The assembly of *D. galeata* was used as a test case and gives an idea of the power of our approach for newly sequenced and assembled *Daphnia* or cladoceran genomes. After performing BLAST analysis for each of the 32903 *D. galeata* transcripts, we observed significant hits (eval < 0; identity ≥ 50%) for 4684 transcripts (Figure 11c). Among them, 4478 *D. galeata* transcripts shared homology with nucleotide sequences, 61 transcripts with protein sequences in the database and 145 transcripts had both a nucleotide and protein sequence hit from the database (Figure 14a).

For *D. galeata*, ~64% of the transcripts were responding to a single stressor (as revealed by our comparative approach), ~20% to two stressors, ~6% to three stressors and less than 4% were associated between 4 and 31 stressors (Figure 14b).

Identifying gene-/transcript-specific stressors is possible when functions of the proteins are known. In our previous attempts to annotate *D. galeata* transcriptome using various tools (Huylmans *et al.* 2016), functions were assigned to about ~66% of the transcripts, and most of them were homologous to *D. pulex* hypothetical protein. Using our comparative genomics approach in the present meta-analysis, we were able to identify transcript specific stressors for 155 transcripts without previous BLAST annotation and 4159 transcripts that were identified to have "hypothetical" functions are now annotated.

Searching the database

Database users should be able to retrieve information on experimentally validated differentially expressed genes in *Daphnia*. A search tool is provided for searching based on the following fields (1) gene ID differentially expressed in reponse to a stressor (2) stressor (3) title of the published research paper (4) authors of the study (5) technique used (6) study species (7) publication year and (8) sequence of the respective gene. The user can query the database using a single field or a combination of multiple fields. The "advanced search" provides a refined way to search the database using several combinations of keywords using logical operators "+" / "-", which represents "AND" / "OR", respectively.

BLAST

To enable searches based on sequence homology and thus extend the use to studies using de novo assemblies or array probes without typical geneIDs, we

implemented a BLAST tool in the database. This tool can be used for similarity-based search of any query sequences, protein and nucleotide. It is to be noted, though, that the database contains only genes responding to stressor and is therefore not an alternative to repositories such as wFLEAbase. The user can submit one or more sequence(s) in fasta format in the search field and set up thresholds for the e-value, among others.

Limitations of the study and recommendations to the community

Although an extensive literature search for retrieving the differentially expressed *Daphnia* genes in response to a stressor was performed, we might have missed studies that did not turn up using the three search strategies followed. However, since the *Daphnia* stressor database is publicly available, data will be added manually and updated on a regular basis. As more data gets published or where our search strategy failed to retrieve the literature, the authors intend to include them in the database for the benefit of the *Daphnia* research community. The researchers also can contact the authors to include their experimentally validated gene expression data.

Because our meta-analysis did not include a re-analysis of all the datasets, we chose to take the inferred changes in gene expression at face value. The caveat is that most cited studies were following standard practice at the time of publication i.e. using state of the art statistical tools and setting similar thresholds for significance, some did not apply multiple testing correction. The results aggregated in our study should therefore be reviewed critically when used to formulate hypotheses and re-analyzed if necessary.

Gene expression in *Daphnia* has been extensively studied in the last decade and the present meta-analysis brought a few weaknesses and strengths to light. *Daphnia* researchers can rely on very extensive data compared to other groups, especially regarding ecologically relevant stressors. The high numbers of stressors and studies constitute a wealth of information and are an invaluable resource for the future. However, the data mining procedure highlighted a few shortcomings, some of which can be easily remedied. Our recommendations for present and future publications are contained in three words: traceable, searchable, and sustainable. First, publications should use traceable IDs to allow for cross referencing and good search results based on gene IDs. Whenever possible, gene IDs already in use in long term online repositories such as GenBank or Uniprot IDs should

be used. Especially qPCR-based approaches were often using non-traceable gene names and including them in results table would be tremendously helpful and even influence the citation record. Second, the results should be searchable, and preference should be given to tables in a universal readable format (csv, tsv, tab delimited txt, spreadsheets to some extent), instead of heatmaps saved as a graphic. Last but not least, sustainability is an issue affecting all publications and becomes increasingly important with the high volume of data produced, associated with publication on one side, and high personal turnover inherent to research on the other side. Both factors are challenging, but solutions already exist in the form of above-mentioned repositories. Such centralized repositories, the Expression Atlas (Petryszak *et al.* 2016) and GEO Profiles (Barrett *et al.* 2013; Edgar *et al.* 2001), gather gene expression data but only for a few selected species. Following these recommendations would allow a better integration of new and older results across laboratories working with *Daphnia*, help formulating new hypothesis, and increase the visibility of the *Daphnia* research community.

Conclusion

Stress response is crucial for an organism's survival in natural habitats. Identifying and functionally annotating ecologically relevant genes is an important aspect of ecotoxicological studies. Signature patterns of gene expression and their responses are often used as markers of environmental stressors. Traditional approaches often subject the genes to laboratory induced stressors and quantify their expression patterns. In the present study, we took advantage of comparative genomics approaches to identify transcript specific stressors in *D. galeata* using data mined from experimentally validated gene expression studies in *Daphnia*. The present database on gene expression studies in *Daphnia* will be useful for the *Daphnia* research community in designing stressor specific experiments and evaluating genes in response to environmental perturbations. For example, approaches such as the overrepresentation analysis presented by Herrmann *et al.* (2017c) and Bowman *et al.* (2018) can be extended to other stressors than temperature. Last, our database will also help interpreting the results of studies on adaptation in natural populations as well as ecological experiments.

Supporting information

The following table is found in the enclosed CD-ROM in the folder "Appendix_Chapter2".

Table A2_T1: List of gene expression studies in *Daphnia*, authors of the paper, year of publication, stressors used, species, experimental technique used in the study.

Acknowledgements

We would like to thank the Volkswagen Stiftung (grant to MC) for financial support. S. Thiemann, M. Herrmann, Y Lu, J Asselman and M. Möst provided useful comments on earlier versions of the database. Rechenzentrum of Universität Hamburg provided assistance for developing the database. This work benefits from and contributes to the *Daphnia* Genomics Consortium.

Chapter 3: An environment-dependent genotype-phenotype association in European *Daphnia galeata*

Suda Parimala Ravindran, Verena Tams and Mathilde Cordellier

Abstract

Environment-dependent phenotypic variation is exhibited to some extent by all organisms. To cope with environmental changes, organisms adapt through a variety of mechanisms such as changes in morphology, physiology, life-history traits or behaviour. Understanding the genetic basis of these phenotypic responses is essential; therefore the identification of candidate genes that mediate phenotypic variation is important. To this aim, we used *Daphnia* as a model organism to understand the transcriptomic basis of phenotypic variation in a predator environment using a transcriptome-wide association approach. Furthermore, we used a gene co-expression network analysis to identify gene clusters correlated to life-history traits. To enhance our understanding of the functional roles of these transcripts, we identified orthologs and paralogs from related species and used ontologies to annotate the candidates of interest. Our association analysis revealed two life-history traits to have a transcriptomic basis in the presence and absence of fish kairomones. Our gene co-expression analysis identified 44 modules, of which one module correlated to the life-history trait, the 'total number of broods'. Thus, our combined use of gene co-expression network and transcriptome-wide association analysis provided a systems-level approach to understand the interplay between the genotype, phenotype and environment in *Daphnia*.

Keywords: *Daphnia* – genotype – phenotype – fish kairomones - GWAS – gene co-expression network - phenotypic plasticity

A similar version of this Chapter was prepared in collaboration with Verena Tams and will be submitted to *Genes, Genomes and Genetics,* a peer-reviewed journal.

Introduction

Global change and its impact on biodiversity is currently in the focus of scientific investigations (reviewed in Beaugrand & Kirby 2018). Natural populations are subject to novel environmental conditions due to climate change, habitat degradation and/or shifts of population ranges thereby expressing new phenotypic characteristics (Grether 2005). When environments impose strong constraints and when adaptive potential exists in populations, selection favors trait values that increase the fitness of individuals in their local habitat. Hence, individuals have better fitness compared to other populations in their local environment and this known as local adaptation (de Villemereuil *et al.* 2018; Kawecki & Ebert 2004).

Phenotypes and genotypes are tightly linked, since a genotype and its environment define the phenotype (e.g. Agrawal 2001; Stearns 1989). Genetic as well as phenotypic variation is crucial for an organism to survive environmental changes and to successfully reproduce and pass their alleles to the next generation. Identifying the genetic basis of local adaptation is critical in addressing the central questions in evolutionary biology (Rausher & Delph 2015). Resolving whether natural selection acts on standing genetic variation or on novel mutations, and identifying the loci contributing to regulatory, coding and structural variation helps to understand the adaptive and speciation processes (Hoban *et al.* 2016).

Phenotypic plasticity is also an important mechanism that helps coping with environmental perturbations (Alberto *et al.* 2013; Charmantier *et al.* 2008). Phenotypic plasticity is the ability of an organism to produce multiple phenotypes from a single genotype depending on the environment (Miner *et al.* 2005) and is considered as a trait by itself (Charmantier *et al.* 2008). Although phenotypic plasticity is advantageous in heterogeneous and/or fast changing habitats, its maintenance is associated with costs (Buskirk & Steiner 2009; DeWitt *et al.* 1998) and sometimes becomes maladaptive (Ghalambor *et al.* 2007; Langerhans & Dewitt 2002). A wide diversity of organisms exhibit plasticity in response to biotic and abiotic factors in their environments (DeWitt & Scheiner 2004; Harvell 1990; Karban 1989; Sultan 2000), leading to changes in behaviour, morphology, physiology and life-history traits. These plastic responses can be expressed

either within the lifespan of a single individual (Young 2003) or across generations (Miner *et al.* 2005).

Individuals in a population show differences in their phenotypic traits, which is influenced by both genetic and environmental sources. Understanding the mechanisms of variation is the key to assess the adaptive potential of a population to changing environments (Fuhrman *et al.* 2018). The most common class of phenotypic variation is quantitative trait variation. Three factors influence the phenotypic trait value of any individual: (i) genetic factors that define heritable differences within an environment (ii) environmental factors that influence the genotypes and (iii) the inherent capability of the phenotypic trait, given that genetic and environmental factors are identical (Ziv 2017). Understanding how each of these factors contributes to variation in quantitative traits remains a challenge.

Examining the genetic architecture of phenotypic traits not only identifies causal mutations but also helps in understanding past and predict future evolutionary processes of adaptation (Ronnegard *et al.* 2016). Genetic variation can be studied at two levels of organization: gene expression and sequence level. Associating the regulatory level genetic variation to phenotypic traits can be done by constructing gene co-expression networks, which identify clusters of co-expressed genes. Often co-expressed genes within one module (cluster) share conserved biological functions revealing their potential genetic pathways (Subramanian *et al.* 2005). The big benefit of gene network analysis lies in the opportunity to correlate the gene expression information to biological information to gather insights of the biological association of genes to certain traits and hence to identify candidate genes. For example, a study on lake whitefish (Filteau *et al.* 2013) used a weighted gene co-expression network analysis to identify gene clusters correlating to three phenotypic traits such as trophic behavior, trophic morphology (gill rakers), and reproduction. Genome wide association studies (GWAS) that link phenotypic traits with genetic data and environment, have been applied extensively in humans (for e.g., Busch *et al.* 2016; Eising *et al.* 2016; Kao *et al.* 2017), animals (e.g., pigs (Duijvesteijn *et al.* 2010), cows (Hayes *et al.* 2009), dogs (Wood *et al.* 2009)) and plants (e.g., rice (Zhao *et al.* 2011), arabidopsis (Atanasov *et al.* 2016), sunflower (Kim & Rieseberg 1999)).

The genotype-environment interaction (GxE) is a common phenomenon describing how a genetic variant has a different phenotypic effect in multiple environments (Smith & Kruglyak 2008). For example, human individuals with sickle cell anemia have a survival advantage in endemic areas of malaria but are at a disadvantage in areas without malaria (Ferreira *et al.* 2011). Recently, biologists have applied genomic data and traditional pedigree information to explain phenotypic differences in life-history traits such as horn shape in soay sheep (Johnston *et al.* 2013), clutch size in collared flycatchers (Husby *et al.* 2015; Ronnegard *et al.* 2016) and Glanville fritillary butterfly (Duplouy *et al.* 2017) where life-history trade-offs may be involved in promoting genetic variation at one or several loci in the species.

Daphnia is a well-established organism for population genetic studies and plays a vital role in the trophic cascade of freshwater ecosystems (e.g. Carpenter *et al.* 2001; Ebert 2005; Sommer *et al.* 2003). This small filter feeding crustacean has become a widely used isogenic model organism in ecology, ecological toxicology and ecological evolution because of its ability to reproduce parthenogenetically. Due to their short generation times and easy handling, *Daphnia* researchers use several individuals of every clonal line in their analyses for measuring both genetic and phenotypic traits. A few studies in *Daphnia* research exist that have associated the phenotypic traits to their genotype by using single nucleotide polymorphisms (SNPs). For example: two previous studies (Henning-Lucass *et al.* 2016; Herrmann *et al.* 2017a), showed the effects of temperature on fitness in *D. galeata*. A study (Schwerin *et al.* 2009) in *D. magna* revealed the effects of temperature on gene expression patterns of several genes. Another study (Bento *et al.* 2017) investigated the association of parasite resistant traits to genotypes in *D. magna*. All the above-mentioned studies use an average trait value for every clonal line and associated them to the genotype.

In the present study, we associate genotypic and phenotypic data of 24 clonal lines of European *Daphnia galeata* by applying two approaches. We performed a genome-wide association (GWA) analysis with the existing life-history trait dataset of Tams et al (*submitted*). This analysis involved 15 individuals for each clonal line in the control and fish environment. We took advantage of a well replicated dataset and associated it to their corresponding transcriptome dataset (Herrmann *et al.* 2017c, Ravindran et al submitted).

We then compared these results to the results from association analysis obtained with the mean phenotypic trait value. Further, we applied a weighted gene co-expression network analysis to understand the phenotype-genotype correlations at the gene network level. Based on these analyses, we addressed the following questions: (1) Which SNPs (genotype) associate to different phenotypic life-history trait in the two different environments? (2) Which gene co-expression modules are correlated to life-history traits in the control environment? We were able to synthesize the two levels of genotype-phenotype associations: (i) genotype (SNPs) – phenotype (life-history traits under control and fish environment) and (ii) gene co-expression modules – phenotype (life-history traits in the control environment). By answering these questions we contributed to the understanding of the interplay between the genotype and phenotype in the absence and presence of fish kairomones in *Daphnia*.

Methods

The genotype (SNP) and phenotype (life-history traits) datasets used in the present study have been described in (Herrmann *et al.* 2017c) and Tams et al (*submitted*), respectively. Variance stabilized normalized reads used in the present study for the gene co-expression network analysis was obtained from Ravindran et al. (*submitted*). Functional annotation of the *D. galeata* transcriptome has been described in Huylmans *et al.* (2016) and Ravindran et al. (*submitted*). We present here a brief overview of methods used for the creation of input datasets and a detailed description of the methods for GWAS and gene co-expression network analysis.

Study organism

The cladoceran *D. galeata* is a widely distributed keystone species in freshwater ecosystems. Their parthenogenic life cycle allows rearing many genetically identical individuals from one genotype. Despite their identical genetic makeup, one clonal line can result in different phenotypes. For all datasets (i.e., genotype (SNP), phenotype, and gene expression) summarized below, we used 24 genotypes (clonal lines) of *D. galeata* from four European lake populations (six clonal lines per population): Greifensee (Switzerland), Jordan reservoir (Czech Republic), Lake Constance (South Germany) and Müggelsee (North

Germany). Clonal lineages were established from dormant eggs extracted from sediment cores, which have been used and described in previous studies (Henning-Lucass *et al.* 2016; Herrmann *et al.* 2017c). They were maintained in lab cultures (18°C, 16h light / 8h dark cycle, food: *Acutodesmus obliquus*, medium: Aachener Daphnien Medium (ADaM) (Klüttgen et al. 1994).

Phenotype datasets and design of life-history experiment

Phenotypic data originates from the life-history experiment investigating the effect of fish kairomones on *D. galeata* (described in Tams et al *submitted)* for a total of 684 experimental individuals (aim: 24 clonal lines x 2 treatments x 15 replicates=720 individuals). Prior to the experiment each clonal line was bred in kairomone-free water (control) and in kairomone water (fish) for two subsequent generations to minimize inter-individual variances (Figure A3_F1). Breeding and experimental phases were conducted at a temperature of 20°C and a 16h light / 8h dark cycle in a brood chamber with a light intensity of 30% (Rumed, Typ 3201D). Experimental individuals (F2) were female neonates of the 3rd to 5th brood. Ten life-history traits were recorded: age at first reproduction (AFR) [day of releasing offspring from brood pouch], numbers of broods per female including numbers of neonates per brood per female ('brood1', 'brood2', 'brood3', 'brood4'), total numbers of neonates per female ('offspring'), total number of broods ('broods'), survival [in days], body length ('size') [in µm] and somatic growth rate ('SGR') [in µm d^{-1}]. The experiment lasted for 14 days for each experimental individual. The experiment revealed a change of life-history trait values when exposed to fish kairomones concordant to previous studies, e.g. early maturation in the presence of fish kairomones. Nevertheless, we found high intraspecific phenotypic variation of life-history traits within each population as well as among the four populations. Further details can be found in Tams et al. (*submitted)*.

In this study we used two phenotypic datasets originating from the above described experiment. First, we used the complete raw dataset with up to 15 individuals as replicates per clonal line (Table A3_T1) Second, we created a dataset containing the means of each clonal line for each of the life-history traits (hereafter, "mean dataset") (Table A3_T2). To avoid confusion with terminology we use the term 'clonal line' for the 24 genotypes and the term 'genotype' for SNPs throughout the manuscript.

Genotype dataset and SNP calling

The SNP calling procedure has already been described in Herrmann et al 2017. Briefly, aligned reads from RNAseq experiment were merged with samtools (Li *et al.* 2009) and realignment around indels was performed using GATK's (DePristo *et al.* 2011) IndelRealigner tool and initial variant calls was made using HaplotypeCaller. Using GenotypeGVCF tool in GATK, samples were jointly genotyped and a single vcf file was obtained. Variants were further filtered using VariantFiltration tool implemented in GATK with the following criteria: (i) clusterWindowSize = 35; (ii) Quality by depth (QD) < 2.0; (iii) Fisher Strand (FS) > 30.0.

To use the SNP data for GWA analysis, we further filtered variants with a minor allele frequency (MAF) of 0.1 to exclude rare variants. Only biallelic sites in the MAF filtered SNP data were considered for further analysis. A total of 155,638 SNPs were used for the association analysis. We used this as input for GWAS using the mean values of the life-history traits (hereafter, "mean dataset") for both univariate and multivariate analysis. However, for GWA using the values per individual for the life-history traits, we artificially inflated the SNP data (hereafter, "inflated dataset") as follows (Figure A3_F2): For every individual from every clonal line, we assumed they contain the same haplotype as *Daphnia* have the ability to reproduce clonally and hence are genetically identical. Therefore, we replicated the haplotypes for every individual in every clonal line in every treatment (24 clonal lines x 15 individuals x 2 treatments = 720) and created two vcf files with SNP information for both the control and fish conditions, separately. For the GxE univariate and multivariate analysis, we combined the SNP information from the two files into a single vcf file.

Association analysis

The program PLINK v.1.07 (Purcell *et al.* 2007) was used to test for association between SNPs (genotype) and life-history traits (phenotype) in the presence (hereafter, fish environment) and absence (hereafter, control environment) of fish kairomones. PLINK was used to perform (a) an univariate approach: where each SNP was tested for association to each of the ten phenotypic traits individually; and (b) a multivariate approach: where each SNP was tested for association to the combination of all ten life-

history traits. Both univariate and multivariate analyses were performed on the "mean dataset", while only an. univariate analysis was performed on the "inflated dataset".

The univariate association was performed using the "assoc" command on each phenotypic trait ("--all-pheno"). The multivariate association was performed using MV-PLINK tool using the "--mult-pheno –mqfam" option. PLINK was further used to test for differences in genotype–phenotype associations between the two environments (genotype-environment interaction) using the "gxe" command.

Both univariate and multivariate analyses were performed for the control and fish environment and the GxE interaction separately. Settings were applied to correct for population stratification in the data set by permuting for 1000 iterations within populations. All p-values were corrected for multiple testing using the "Bonferroni correction" method in R (R Development Core Team 2008). The $-\log_{10}$ *p*-values were calculated on adjusted *p*-value and visualized using Manhattan plots in R. A SNP was said to be associated to a phenotypic trait if it had a $-\log_{10}$ *p*-value of 1.5.

Gene co-expression network analysis – Linking gene co-expression and life-history traits (under control conditions)

The terminology of weighted gene co-expression network analysis has been described previously (Langfelder & Horvath 2008). The gene co-expression network analysis was based on variance stabilized read counts obtained from HTSeq data used in the R package 'DESeq2' (Love *et al.* 2014) by Ravindran et al. (*submitted*) to investigate differential gene expression at population level between the four European *D. galeata* populations in control environment. We applied a weighted gene co-expression network analysis (WGCNA) by using the R package 'WGCNA' v. 1.6.1 to find putative pathways from the highly correlated genes clustered in modules (Langfelder & Horvath 2008).

First, an automatic signed weighted, single gene co-expression network was constructed on a workstation with the R environment v.3.2.3 while subsequent analysis was performed in the R environment v.3.4.2 (R Core Team 2017). Second, gene co-expression modules were identified using the WGCNA Topological Overlap Matrices (TOM) with a soft cut-off threshold of 8. Module eigengenes (ME) representing the average gene expression of their module were calculated and used to infer correlation with life-history

traits following a resampling procedure outline as below. Finally, the most interconnected genes per module, known as 'hub-genes' were identified.

Module eigengene-trait correlation

Gene co-expression modules were correlated to phenotypic life-history trait data. In the gene expression analysis, we had three biological replicates per clonal line, while we had one mean value for every life-history trait measured per clonal line. To perform correlation analysis in WGCNA, we had to assign the same mean trait value to all three biological replicates resulting in potential false or inflated correlations (pseudoreplication). To avoid this artefact, we randomly resampled the available individual trait values for each life-history trait in every clonal line, to obtain one "unique" mean trait value per replicate and clonal line. For example, we had trait values for 15 individuals for the trait 'broods' in the clonal line J2 (Table A3_T1). In the first resampling step, we randomly picked 75% of the individuals to calculate a mean. The process was repeated twice to obtain 3 randomized mean values for this life-history trait per clonal line. This step was repeated for every trait value in every clonal line. Finally, the correlation of module eigengenes and the resampled life-history trait mean values was calculated. This whole procedure of resampling to calculate randomized means and their correlation to module eigengenes was repeated 10,000 times to verify the robustness of ME-trait correlations. We then counted the number of observations ME-trait, where the correlation value was above an arbitrary threshold of ±0.5. ME-trait correlations were considered robust if it occurred in more than 95% of the iterations (R script available upon request). Significant ME-trait correlations resulted in transcript sets if interest which were extracted to investigate their biological significance. In addition, hub-genes, which represent the most interconnected gene per module, were identified and their biological importance was explored.

Functional annotation

For every SNP/transcript associated to a life-history trait in both GWA and gene co-expression network analysis, we assigned Gene Ontology (GO) terms using annotations from Huylmans *et al.* (2016) and performed an enrichment analysis with "topGO" (Alexa & Rahnenfuhrer 2016). Additionally, we also identified orthologs and paralogs for the

transcripts associated to a phenotypic trait using orthoMCL data from Huylmans *et al.* (2016). To enhance our understanding of the ecological role of transcripts associated to life-history traits, we performed a BLAST analysis on the *Daphnia* stressor database (Ravindran *et al*, manuscript in preparation). This enabled us to identify stressors for the candidate transcripts of interest from our GWAS and gene co-expression network analysis.

Results and discussion

Genotype-phenotype associations (GWAS analysis)

(a) "Inflated" dataset: univariate analysis

We wanted to take advantage of our well replicated life-history measurements and performed an association analysis considering every individual, rather than phenotypic mean values and inflated the genotype data. Our analysis revealed associations between the genotype and phenotype for all traits (except survival) in both control and fish environments (Table A3_T3). In the GxE analysis, we found seven traits ('brood2', 'brood3', 'AFR', 'broods', 'survival', 'size' and 'SGR') to be associated with the genotype.

Although we found a large number of SNPs to be associated to life-history traits in the inflated dataset, we could not differentiate between the true positive and false positive associations. One reason is the lack of available literature for comparisons that use such experimental setups and inflated genotype calls. Second, existing GWAS tools are not capable of handling replicate data as such for organisms like *Daphnia*. Since this kind of GWAS setup leads to excessive background noise, we could not infer the true associations. Therefore we used the "mean" dataset for further analysis.

(b) "Mean dataset": univariate analysis

While taking the mean values of replicates for each clonal line, we observed associations between genotypes and two life-history traits, namely brood3 and brood4 in the univariate analysis (Table 3). In the fish environment, the phenotypic trait brood3 was associated with three SNPs (in three different transcripts). Brood4 had one SNP in the fish environment and two SNPs (in two transcripts) associated in the control environment. Significant GXE interactions across the two environments were found for one phenotypic

trait, namely 'offspring', associated with two SNPs (from one transcript). No other associations were found for the other life-history traits.

We found no direct evidence of genotype-phenotype associations for most of the life-history traits except for brood3 and brood4. The genotype-phenotype associations for brood3 and brood4 are promising candidates for true associations. To verify these associations a subsequent investigation should test the candidate SNPs for their biological relevance for the trait. The observed significant GxE effect for the trait offspring could be best explained by the significant associations of SNPs to brood3 and brood4 since these life-history traits are not independent. The trait offspring describes the total number of offspring per female and thus includes the total number of offspring for the first, second, third and fourth brood (brood1 to brood4). This correlation of life-history traits might have biased the statistical association of genotype and phenotype and led to a false positive association in the GxE analysis.

Table 3: Number of significant SNPs and corresponding transcript associations of each life-history trait in control, fish and GxE interaction "mean" dataset.

	Control: $-\log_{10}P \geq 1.5$	Fish: $-\log_{10}P \geq 1.5$	GxE: $-\log_{10}P \geq 1.5$
brood1_snps	0	0	0
brood1_transcripts	0	0	0
brood2_snps	0	0	0
brood2_transcripts	0	0	0
brood3_snps	0	3	0
brood3_transcripts	0	3	0
brood4_snps	2	1	0
brood4_transcripts	2	1	0
afr_snps	0	0	0
afr_transcripts	0	0	0
broods_snps	0	0	0
broods_transcripts	0	0	0
offspring_snps	0	0	2
offspring_transcripts	0	0	1
survival_snps	0	0	0
survival_transcripts	0	0	0
length_snps	0	0	0
length_transcripts	0	0	0
sgr_snps	0	0	0
sgr_transcripts	0	0	0

Although significant phenotypic differences were observed in the life-history trait experiment for all the four populations (Tams *et al* submitted), our present GWAS results did not allow pinpointing the associated sequence polymorphism. This implies that the traits are phenotypically plastic (Ayrinhac *et al.* 2004) and thus not necessarily under immediate selection (Merila & Hendry 2014). The number of associated SNPs is concordant to previous studies such as the one associating clutch size in a bird to 50,000 SNPs identifying one significant SNP association (Husby *et al.* 2015). Low levels of genetic variation in populations can also prevent an adaptive response as seen for example in *Drosophila birchii* (Hoffmann et al. 2003).

There are three possible explanations that might describe the lack of association between the genotype and phenotypic trait. A first explanation is that our failure to find SNPs associated to the life-history traits does not imply a lack of genetic variation among the populations, as we observed significant genetic variance at the sequence-level among the studied *D. galeata* populations (Ravindran *et al.,* submitted). GWAS studies are most powerful for phenotypic traits with strong effects that are often caused by one or few loci, with a minor sensitivity to environmental variation (Routtu *et al.* 2014), which is rarely the case. The absence of significant associations between the genotype and phenotypic traits allows for speculation about the underlying genetic variation. A preliminary speculation is that although we had a good number of replicates for each genotype, in the present analysis ("mean dataset"), we had taken only one mean value for each genotype. However, further analysis is required to test the effects of multiple SNPs to one phenotypic trait. For example, a study by Routtu *et al.* (2014) in *D. magna* reported that 8.2% genetic variation was explained by the significant QTLs detected for phototactic behaviour measured in the absence and presence of fish kairomones, stems from multiple genes with small effects. Similarly, another study by De Meester (1991) in *D. magna,* reported that the genetic variation underlying phototactic trait is largely additive in nature. Thus, the phenotypic traits might reflect a pattern of additive, polygenic inheritance.

A second possible explanation for the absence of genotype-phenotype associations might be due to observed heterozygosity measures. Most of the clonal lines (19 out 24) in the sampled populations showed higher observed heterozygosity than the expected levels (Ravindran *et al.,* submitted). This pattern hypothetically makes the clonal lines less plastic

as higher heterozygosity levels may help to buffer environmental influences on the phenotype (Pigliucci 2005).

No significant associations between the phenotype and the transcriptome-based SNPs could also imply that the genetic basis of phenotypic variation is not predominantly on the coding sequences, but rather on the non-coding regions of the genome. Such associations between phenotypic traits and polymorphism in non-coding regions has been reported by McKown *et al.* (2014) in *Populus trichocarpa,* where 152 out of 275 identified associated polymorphisms were in non-coding regions. To test the role of non-coding regions and their association to the life-history traits in *D. galeata*, genomic data is required.

(c) "Mean" dataset: multivariate analysis

As our univariate approach showed weak association signals between the genotype and life-history traits, we performed a multivariate testing. Multivariate testing has been shown to be more powerful compared to univariate analysis (Galesloot *et al.* 2014). Therefore, we assessed the effect of one SNP on all phenotypic traits combined, thus taking the interdependence/correlation of the life-history traits into account.

We identified 38 SNPs (in 24 transcripts) to be significantly associated with all life-history traits in the control environment (Table 4). However, no SNPs were significantly associated to all phenotypic traits in the fish environment. Our multivariate GxE analysis showed 51 SNPs (in 40 transcripts) to be associated to all life-history traits.

Table 4: Number of significant SNP and corresponding transcript associations of multivariate analysis for the control, fish and GxE interaction "mean" dataset.

	Control: $-\log_{10}P \geq 1.5$	Fish: $-\log_{10}P \geq 1.5$	GxE: $-\log_{10}P \geq 1.5$
All_SNPs	38	0	51
All_Transcripts	24	0	40

In general, the association of SNPs to complex traits has been reported to be successful (Galesloot et al 2014). Since PLINK is known to perform better compared to other tools available for multivariate genome-wide association studies (Galesloot *et al.* 2014), we do not expect a statistical/analytical bias per se. The number of identified

genotype-phenotype associations in the control environment as well as for the genotype-environment interaction (GxE) is concordant to a previous study in wing shape of *D. melanogaster* that identified 139 SNPs to be associated to the phenotype (Pitchers et al 2017). The phenotypic traits used for analysis are directly or indirectly correlated with reproduction. The number of SNPs associated to life-history traits in the multivariate GxE might exist because the trait values in the control environment drive the statistical associations even though there is a lack of association in the fish environment. Further analysis is required to test this interpretation such as knock-down studies on the specific transcripts that are inferred to be associated to the life-history traits. Furthermore, we cannot rule out other possibilities of associations at the non-coding regions or epigenetic level (Tak & Farnham 2015).

Gene co-expression network analysis – Linking gene co-expression and life-history traits (WGCNA analysis)

The single network construction resulted in 44 modules of co-expressed transcripts in the control environment (Figure A3_F3). Most transcripts were assigned to the modules 'turquoise', 'blue', 'brown' and 'yellow'. The 'grey' module is the largest and includes all transcripts which could not have been assigned to any module; it comprises 22% of all transcripts (n=7297). For each module the 'hub-gene' i.e., the most highly interconnected gene within a gene co-expression module, was identified. To assess the biological meaning of modules we correlated life-history trait information to the module eigengenes (ME). Only one module, 'darkorange' was significantly ($p \leq 0.05$) correlated (in 9,782 out of 10,000 correlations) to one life-history trait, the 'broods'. A detailed overview of modules, number of transcripts and hub-genes are listed in Table A3_T4.

Functional annotation

Our integrative approach revealed genotype-phenotype associations at sequence and regulatory level. A univariate analysis identified a total of eight SNPs to be associated to three life-history traits (4 SNPs in fish, 2 SNPs in control environment and 2 SNPs in GxE), while the multivariate analysis revealed 38 SNPs in the control environment, no SNPs in the fish environment and 51 SNPs by the GxE interaction. A gene co-expression analysis

revealed an association to one co-expression module (darkorange) with 85 annotated transcripts. In summary, we identified a list of 157 candidate transcripts being involved in phenotypic variation in *D. galeata*.

Gene Ontology analysis

Gene Ontology (GO) terms were assigned to transcripts identified in the univariate and multivariate GWA analysis, in addition to hub-genes and transcripts of the darkorange module which correlated to the trait broods in the WGCNA analysis. In total, GO terms were assigned to 68 transcripts (Table A3_T5; 44 transcripts in GWA and 24 transcripts in WGCNA) and to 15 out of the 44 hub-genes (Table A3_T4). GO terms identified in the GWAS analyses were enriched for spermatogenesis and other metabolic processes; and those identified in the WGCNA analyses were enriched for metabolic processes (Table A3_T6). There were 18 GO terms assigned to the hub-genes and included functions for enzyme activities, binding and transport activities which are important for general metabolic processes.

Surprisingly, we found GO terms of the GWA analysis to be enriched for spermatogenesis. There is no sound explanation to this observation as only parthenogenetically reproduced females were used in the experiment. We cannot exclude that a statistical bias exists because only ~31% the *D. galatea* reference transcriptome has GO annotations.

The 17 enriched GO terms (Table A3_T6) assigned to the transcripts of the darkorange module from the WGCNA analysis were for enzymatic activities and metabolic processes. Our GO analysis for the identified candidate transcripts suggests that there might be a hierarchal activation of general stress response mechanisms at both the metabolic and cellular level. Another study subjected *D. magna* to several biotic and abiotic factors (Orsini *et al.* 2017) and revealed a general stress response rather than ontologies specific to inducible defenses or other environmental perturbations. Additionally, the gene set enrichment analysis emphasizes the need of further functional annotations for the existing *Daphnia* genomes to improve biological valid conclusions. By now most of the transcripts are without annotation. Thus, we cannot exclude a bias in the gene set enrichment analysis due to lack of additional GO terms.

Comparative genomics

To identify the species specificity of the candidate transcripts from GWAS and WGCNA approaches, we performed an orthoMCL analysis. Out of the 157 candidate transcripts, 67 transcripts obtained from GWAS and 53 transcripts obtained from WGCNA had an orthogroup assigned. One orthogroup, the ORTHO_ALL24 cluster was common to both GWAS and WGCNA approaches, which was involved in protein binding process (GO:0005515). Our orthoMCL analysis shed light into the paralogs and orthologs for the candidate transcripts of interest. Although our GWAS and WGCNA analysis revealed only a few *D. galeata* transcripts to be associated to the life-history traits, the paralogs identified through the orthoMCL analysis might also be associated to the phenotypic trait. The orthologs and paralogs identified for the candidate transcripts in the present study can act as a basis for future research in association studies in *Daphnia* and insects.

Transcript-specific stressors

Identifying transcript specific stressors is another way of looking into the functional aspects of a transcript. In the present study, we identified stressors for 10 candidate transcripts in GWAS analysis and 11 transcripts in WGCNA analysis (Table A3_T5). Although we did not find any of the candidate transcripts to have a homologous gene that responds to kairomones, we found a few candidate transcripts to have a homologous gene that expresses itself to several biotic and abiotic stressors. This could imply that our candidate transcripts identified in the present study could be a part of the general stress response, in addition to be affected by predator kairomones.

Depending on the genes or transcripts of interest, further research can be emphasized on these candidates that have an identified stressor in literature (Figure 15). However, the number of candidate transcripts with an associated stressor is very few (~13%) and hence we cannot infer the stress responses more precisely. This highlights the need for improved annotations in *Daphnia*.

Limitations and conclusions

In this study, we explored the association of phenotype, genotype and environment in European *Daphnia galeata* emphasizing the complexity of their interactions. The

present study revealed very few of the life-history traits to have a transcriptomic basis and also brought to light some shortcomings. First, an appropriate genome-wide association approach is missing to account for the clonal nature of *Daphnia*. We would have gained more information and statistical power by using the complete phenotypic dataset of individuals (n= ~700) rather than phenotypic means per clonal line (n= 24). Second, although we found very little evidence at the transcriptome level, we cannot exclude the role of non-coding regions in shaping phenotypic plasticity. Therefore, genomic data for *D. galeata* would have helped in testing this hypothesis. Genomic information can further enhance our understanding of the interplay between genotype, phenotype and environment. In addition, an easy access to annotation information for *Daphnia* would help to identify biologically meaningful transcripts. Third, further analysis is required to test the effects of multiple SNPs affecting one phenotypic trait, which can help to identify the role of polygenic trait inheritance. Fourth, to better understand the influence of predation risk simulated by the presence of fish kairomones on *Daphnia* life-history traits, gene expression profiles are needed for all 24 clonal lines. These gene expression profiles would provide new insights of this study and would allow the application of a differential gene co-expression network analysis between the two gene co-expression networks (control vs. fish kairomones). Furthermore, the availability of gene expression profiles for both control and fish kairomones data would further reveal biologically significant pathways and hence candidate transcripts. Overall, the identification of biologically significant transcripts being involved in predator-induced responses in *Daphnia* provided a valuable source for further investigations of the environment-dependent genotype-phenotype relationships in *Daphnia*.

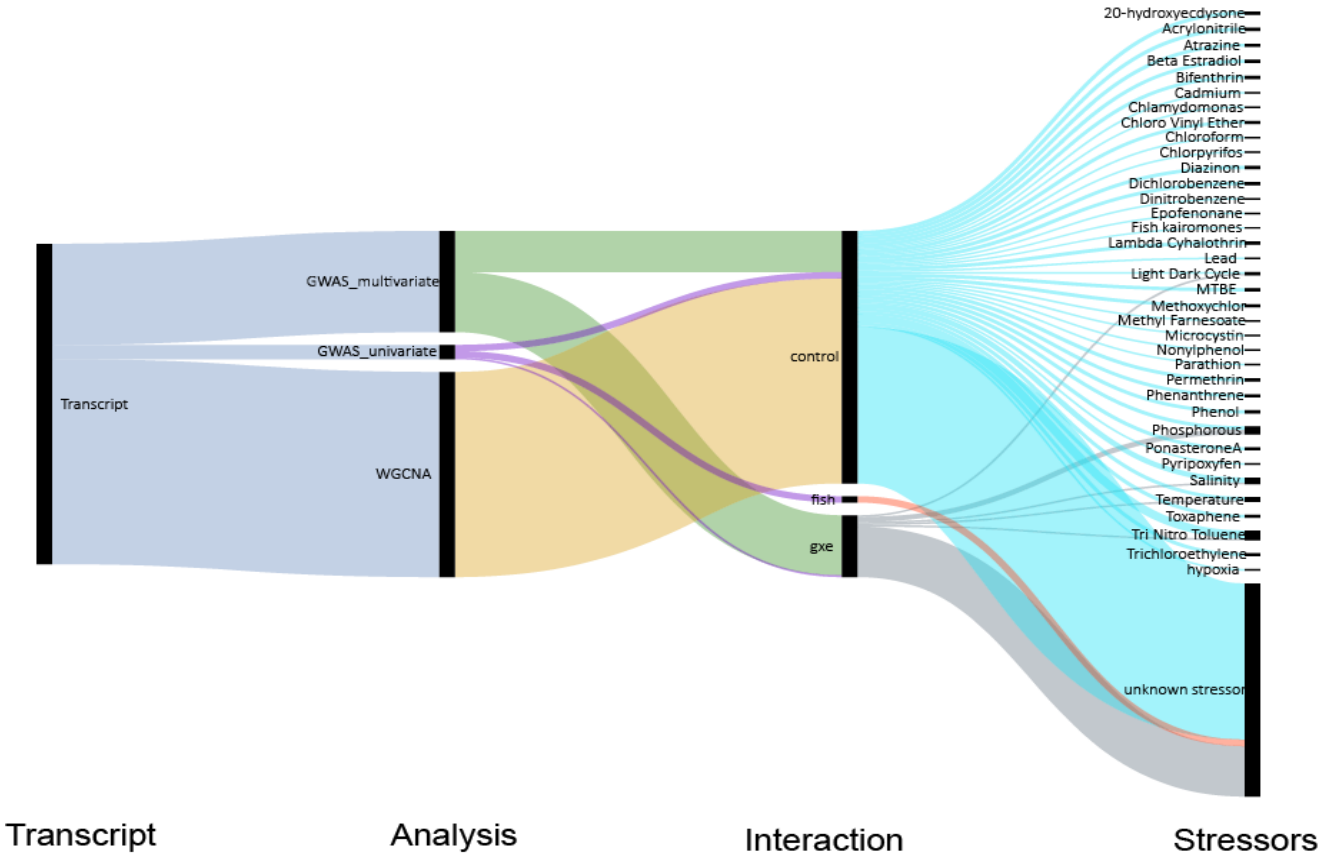


Figure 15: Flow diagram representing the proportion of candidate transcripts as identified in GWAS and WGCNA and their associated stressors.

Each rectangle bar is called a 'node' and each vertical group of nodes is called a 'step'. The colored areas linking the nodes are called 'flows'. The step 'Transcript' contains candidate transcripts as identified in GWAS and WGCNA. The step 'Analysis' contains three nodes: 'GWAS_multivariate', 'GWAS_univariate' and 'WGCNA' representing the analysis from which the candidate transcripts were obtained. The step 'Interaction' contains three nodes: 'control', 'fish' and 'gxe' representing the control, fish environments and gxe interactions. The step 'Stressors' represents the identified stressor for each transcript based on sequence similarity from *Daphnia* stressor database.

Acknowledgments

The authors would like to thank Dr. Maike Herrmann for her useful inputs on working with PLINK.

Data Accessibility

SNP data used as input for GWAS analysis will be archived in Dryad online repository and accession numbers will be provided.

Supporting information

The following tables and figures are found in the "Appendix for Chapter 3" section.

Table A3_T3: GWAS results of "inflated" dataset in control and fish environments as well as GxE interaction.

Table A3_T5: Identification of transcript-specific stressors, orthologs and paralogs for the candidate transcripts of interest as identified in the univariate and multivariate GWAS and WGCNA. A total of 156 candidate transcripts are listed. 'orthogroup' refers to the orthoMCL cluster the transcript is assigned to, 'dgal' = number of *D. galeata* transcripts present in the assigned orthoMCL cluster. 'dpul' = number of D. *pulex* genes present in the assigned orthoMCL cluster. 'dmag' = number of *D. magna* genes present in the assigned orthoMCL cluster. 'dme' = number of *Drosophila melanogaster* genes present in the assigned orthoMCL cluster. 'nvi' = number of *Nasonia vitripennis* genes present in the assigned orthoMCL cluster. 'hit' refers to a transcript that is significantly similar in *Daphnia* stressor database to the candidate transcript of interest. 'identity_percent' refers to BLAST identity percentage corresponding to the *Daphnia* stressor database hit.

Figure A3_F1: Breeding design of life-history experiment in the absence or presence of fish kairomones (Tams et al *submitted*). Each clonal line was bred in kairomone-free water (control environment) and in kairomone water (fish environment) for two subsequent generations (F0 & F1). Neonates from 3rd to 5th brood were used to start a new generation. Life-history traits of experimental individuals (F2) were measured for 14 days (t14). Neonates were preserved in ethanol at the beginning of the experiment (t0) and experimental individuals at the end of the experiment (t14) were used to measure the trait, 'length').

Figure A3_F2: A visual representation of how the "inflated dataset" of SNPs was created for GWA analysis.

The following tables and figure are found in the enclosed CD-ROM in the folder "Appendix_Chapter3".

Table A3_T1: Raw life-history trait data used as input for GWA analysis in the control and fish environments. The life-history traits are 'total number of offsprings per brood (1st

brood= 'brood1', 2nd brood= 'brood2', 3rd brood= 'brood3' and 4th brood= 'brood4'), age at first reproduction ('AFR' i.e., day of releasing neonates from brood pouch), total number of broods ('broods'), total number of offspring ('offspring'), body size ('length', in µm), 'survival' and somatic growth rate ('SGR', µm/day).

Table A3_T2: Mean values of the life-history trait data used as input for GWAS analysis in the control and fish environments. The life-history traits are 'total number of offsprings per brood (1st brood= 'brood1', 2nd brood= 'brood2', 3rd brood= 'brood3' and 4th brood= 'brood4'), age at first reproduction ('AFR' i.e., day of releasing neonates from brood pouch), total number of broods ('broods'), total number of offspring ('offspring'), body size ('length', in µm), 'survival' and somatic growth rate ('SGR', µm/day).

Table A3_T4: Overview of gene co-expression modules in *Daphnia galeata* in control environment from WGCNA. The table summarizes module color, total number of transcripts per module, the name of the most inter-connected gene (hub-gene), as well as Gene Ontology (GO) IDs and classes. The module 'grey' contains all co-expressed genes which were not assigned to a co-expression module.

Table A3_T6a-f: List of GO enrichment for candidate transcripts of interest as identified in GWA analysis (S6a-c) and WGCNA (S6d-f). The three ontologies are: molecular function (MF), cellular component (CC) and biological processes (BP).

Figure A3_F3: Cluster dendrogram of D*aphnia galeata* transcripts obtained from WGCNA analysis. Dissimilarity based on topological overlap matrices (TOM). Additional assignments are the module colors, the gene significances (GS) for the trait clone (clonal line) and broods (total number of broods). Red and blue colors indicate a positive and negative correlation of the module with the respective trait. Darker hues indicate higher correlation values.

D. General discussion

In my thesis, I focused on assessing the transcriptomic basis of local adaptation, linking the genotype and phenotypic traits, and functional annotation of *Daphnia galeata*. To this aim, I used the *D. galeata* transcriptome to identify candidate transcripts that are under local adaptation at the regulatory and sequence level and annotated their functions (**Chapter 1**). The results from this analysis delivered a mixed message i.e., the divergence at the sequence level differed from the divergence at the regulatory level by several orders of magnitude. Although, *D. galeata* transcripts were functionally annotated using different approaches, many transcripts were "hypothetical" or "unknown" in function. Therefore, to further enhance the ecological and functional role of these "hypothetical" transcripts, I used a data mining approach to identify genes that are differentially expressed under different stressors from several previous studies (**Chapter 2**). Using a comparative genomics approach, I was able to improve the existing functional annotation of *D. galeata* transcripts. The data mined from literature was developed into a database called the "*Daphnia* stressor database". Lastly, I concentrated on revealing the associations between the genotype and phenotype in *D. galeata* populations. I utilized the life-history trait data (ten traits) obtained from the same clonal lines of *D. galeata* in control and fish-kairomone environments from another experiment. Using the SNP data from **Chapter 1**, I performed an association analysis to the ten life-history traits to assess the environment dependent genotype-phenotype relationships of *D. galeata* (**Chapter 3**). Results from the association analysis revealed that there is an association between SNPs and one life-history trait in the control environment and two traits in fish-kairomone environment. I also revealed genotype-environment (GxE) interaction for one life-history trait. Furthermore, a weighted gene co-expression network analysis using the expression profiles from **Chapter 1** and phenotype data revealed gene clusters correlated to one life-history trait. I also identified orthologs and paralogs from related species and used gene ontologies to annotate the candidate transcripts as revealed from the association and gene co-expression network analysis. Thus, I conclude that although there are several candidates identified for local adaptation at the regulatory and sequence level, there is very little evidence for transcriptomic basis for the observed variation in the life-history traits. Annotating *D.*

galeata transcripts using a comparative approach further augmented the understating of the ecological role of *Daphnia*. These results convey important implications for the aims of my thesis: (1) What is the role of natural selection and genetic drift on gene expression and sequence variation within and between four lake populations of *D. galeata*? (2) What is the functional role of the *D. galeata* transcripts? (3) What conclusions can be drawn from the genotype-phenotype relationships of *D. galeata* in the absence and presence of a predator kairomone?

D.1 Local adaptation in Daphnia galeata

The results presented in **Chapter 1** of my thesis which was to distinguish neutral from adaptive evolutionary processes at the gene expression and DNA sequence level, delivered a mixed message. At the sequence level, I observed patterns likely due to initial founder effects, bottleneck events, genetic drift and divergent selection. At the gene expression level, I observed high variation within *D. galeata* populations and less variation between populations indicating balancing and directional selection. In addition to the interpretations discussed in **Chapter 1**, here I discuss on other possible explanations for my observations.

D.1a Genetic drift

Genetic variation among populations may result from natural selection acting on adaptive variation or from stochastic events such as founder effects and genetic drift during initial colonization or during bottleneck events (De Meester 1996a). During stochastic events such as genetic drift, founder events and bottleneck events, the allele frequency increases than they were previously and can reach even higher frequencies due to strong genetic drift, which occurs while the population size is still small (Slatkin 2004). When individuals colonize new habitats, they are subjected to novel selection pressures and develop new adaptations which differentiates them from their ancestors (Matute 2013). When colonization occurs by a small number of founders that are in a population bottleneck, the genetic composition of the colonizing population differs from the ancestral population due to genetic drift. This process might hinder adaptation and evolution of the founding populations (Matute 2013). When the founding populations have a lower

population size, the new populations can be confounded by the effects of inbreeding. For example, high levels of inbreeding increase the additive genetic variance and leads to decanalization, which can overcome genetic variation (Hill *et al*. 2006; Robertson 1952; Wang *et al*. 1998; Willi *et al*. 2007; Willis & Orr 1993). The loss of genetic variation caused by smaller population sizes has a variety of harmful effects on growth, development and survival rates, thereby reducing the evolutionary potential of these populations (Allendorf 1986). When populations are founded by a few individuals, the recovering populations might suffer from a low or even a lack of genetic variation that is usually associated with a bottleneck, which can elevate inbreeding depression and the probability of fixation for deleterious mutations (Heather *et al*. 2003). Therefore, the pattern of genetic drift, initial founder effects and bottleneck events that I observe at the sequence level in **Chapter 1** of my thesis was not very surprising as evidence exist that the lakes of Greifensee (Straile 1998), Müggelsee (Okun *et al*. 2005) and Constance (Brede *et al*. 2009) have undergone historical bottleneck events. Furthermore, genetic diversity may be affected through drift because *Daphnia* populations have an innate demographic instability with repeated bottleneck events (Östman 2011). Such conditions might rapidly improve with even minor levels of immigration (Heather *et al*. 2003). However, further investigations are required into the sources of the colonists in order to understand the processes affecting the aquatic community.

D.1b Natural selection

Natural selection was the most obvious explanation for several candidate transcripts I identified in **Chapter 1** of my thesis at both the expression and sequence level. Diversifying selection was the obvious explanation for population divergence at the sequence level along with other factors explained in **Chapter 1**. At the regulatory level, balancing selection was the explanation for candidates identified in **Chapter 1** of my thesis. Balancing selection helps to preserve the coexistence of multiple adaptive variants for any given trait within a population. This phenomenon has important implications for the genetic architecture of adaptive traits, leading to the following predictions (Llaurens *et al*. 2017). First, when a single trait is controlled by multiple loci, the adaptive combinations of alleles is broken down due to recombination (e.g. Charlesworth & Charlesworth 1975).

Second, the proportion of heterozygotes in a population increases, thereby leading to evolution of dominance at loci under balancing selection (Bourguet 1999). This might allow the accumulation of deleterious mutations rarely expressed in heterozygous genotypes (van Oosterhout 2009). However, further analysis and genomic data along with linkage maps would be helpful to unravel the mechanisms of speciation in *D. galeata* populations.

D.1c Dispersal and gene flow

Although genetic drift and selection played an important role in explaining the population divergence at the regulatory and sequence level (**Chapter 1**), the involvement of gene flow and dispersal in the populations cannot be completely excluded. Variability in environmental conditions among habitats may lead to local adaptation of populations, and this process is counterbalanced by gene flow between populations (Slatkin 1985). Local adaptation is hypothesized to play an important role in shaping population divergence especially in aquatic species with high rates of dispersal, rapid population growth and capability to produce resting eggs, like *Daphnia* (De Meester *et al*. 2016). *Daphnia* often rapidly colonize new ponds (Fryer 1985) and their resting eggs are resistant to freezing (Wood 1933), drying (Davison 1969), mechanical stress, and digestive enzymes (Mellors 1975; Proctor 1964). Thus, the dispersal capabilities of *Daphnia* are believed to be reasonably high, which can drive gene flow to homogenize genotype frequencies among *Daphnia* populations. However, substantial genetic differentiation among *Daphnia* populations has frequently been reported and has been observed for ecologically relevant traits (Ebert 2005). This apparent ambiguity between genetic differentiation and dispersal potential of *Daphnia* could be due to the stochastic effects associated with clonal selection, which can enhance genetic differentiation (Vanoverbeke & De Meester 1997a). However, the possibility of disparity in effective gene flow cannot be excluded. Because of their apomictic behavior, *Daphnia* have the ability to adapt genetically to their local environment, which promotes an 'edge' for any genetic variation in quantitative traits (Lynch 1983). Hence, local adaptation may be attained rapidly and then further assisted by competitive exclusion of less adapted invading genotypes (Korpelainen 1986), outbreeding

depression (De Meester 1993; De Meester 1996b) and monopolization effects (De Meester *et al*. 2002; Louette *et al*. 2007; Munoz *et al*. 2016; Ortells *et al*. 2013).

According to the Monopolization Hypothesis (De Meester *et al*. 2002), local adaptation along with founder and priority effects prevents gene flow and causes an increase in genetic differentiation, as often observed in *Daphnia* populations (Thielsch *et al*. 2009; Vanoverbeke & De Meester 1997a). Boileau *et al*. (1992) in a theoretical study on zooplankton which can colonize new habitats from a few propagules and have rapid population growth, showed that even with higher gene flow rates in populations, allele frequencies established during initial founder events are resistant to decomposition for several generations. The initial colonizers tend to increase rapidly, occupying the entire habitat in a few generations and therefore subsequent migrants are strongly diluted leading to negligible impact on allele frequencies (De Meester *et al*. 2002). Such persistent founder effects reflect priority effects in which there is a conflict in the number of residents and immigrants, leading to lower random increase of migrants due to drift alone (De Meester *et al*. 2002). For example, strong founding events have been in observed in *Daphnia pulex* (Snæbjörn 2001) and rotifers (Gómez *et al*. 2002). Priority effects are common in several invertebrates found in ponds and lakes. This strong priority is because of the inherent growth capacity of organisms like cladocerans, rotifers and large branchiopods. Furthermore, the presence of resting propagule banks acts as a buffer against new invaders, thereby population size is maintained at high levels and increased generation times, which therefore takes a longer time to reach the migration-drift equilibrium (De Meester *et al*. 2002). The presence of continuous founder effects hypothesizes that with increasing habitat size, genetic differentiation between populations decrease. For example, a study by Naihong *et al*. (2000) between *Artemia* populations inhabiting large lakes showed that the observed low levels of genetic differentiation can be explained with increasing habitat size.

On the contrary, where population bottlenecks and extinctions are common in very small *Daphnia* populations and are continuously recolonized by a single individual, inbreeding may occur, leading to strong fitness during parthenogenesis and sexual recombination in the immigrant genotype (Ebert *et al*. 2002). The study on *Daphnia* populations in Finland by Ebert *et al*. (2002), showed that sexual recombination as an

advantage for the immigrant genotypes, leading to inbreeding. However, most of the other *Daphnia* studies reported in literature do not follow this pattern, instead local adaptation (i.e. the average fitness of the resident populations compared to the immigrant individuals) plays a major role in decreasing colonization events as evidenced by strong outbreeding depression (De Meester *et al.* 2002). It is likely that immigrant *Daphnia* individuals are affected by selective sweeps and genes that occur by chance are better adapted to a specific habitat is low, however, this probability will be higher in very small habitats or newly founded individuals or bottlenecked populations. Therefore, to assess the relationships between inbreeding, habitat size and local adaptation, more detailed studies are needed to determine the genetic structure of *Daphnia* populations in smaller habitats (Berg 2001).

D.2 Genotype-phenotype associations

The number of phenotypic traits that had a transcriptomic basis of variation in *D. galeata* was very low in the univariate genome-wide association analysis in both control and fish environments (**Chapter 3**). This lack of association implied that several traits are phenotypically plastic by themselves. A multivariate genome-wide association analysis provided hints at the roles of multiple traits associated to a SNP. This implied that the SNPs affect not just one trait, but the combination of multiple traits. Our results were consistent with several previous studies for example in field mice, butterflies and stickleback fish have identified a small number of genes that have an effect on the morphological characteristics (Brennan *et al.* 2018). However, further studies are required to draw conclusions about the genotype-phenotype associations in *Daphnia*. For example, the experimental time used for the life-history trait data in the genome-wide association study (**Chapter 3**) was 14 days. However, exposing the daphniids to kairomones for shorter or longer periods of time might have detected signals of associations between the genotype and phenotype. Therefore, to find evidence of adaptive genetic changes, both genetic and phenotype data over time and fitness measurements through a putatively adaptive allele is required (Hansen *et al.* 2012).

On the other hand, it would be useful to evaluate the contribution of multiple SNPs to one phenotypic trait. The levels of transcriptomic variation for quantitative traits and

molecular markers are expected to differ theoretically, because polygenic traits have higher mutation rates in general and greater exposure to selection pressures (Morgan *et al.* 2007). For several adaptive challenges, phenotypic traits might be highly polygenic and multiple traits may contribute to the raise of a complex adaptive phenotype. This is true for organisms in aquatic habitats as they are subject to several environmental features such as salinity, pH, predators, pathogens, etc., which affects their fitness (Brennan *et al.* 2018).

Although GWAS is a widely used method to identify the relationships between the genotype and phenotypic variation, it is inefficient to identify variation associated to polygenic traits or vice-versa. Instead, genome-wide scans for selection signatures can be used for discovering genomic regions that underlie local adaptation. However, selection scans do not provide evidences for specific genotypic variation with phenotypes. Therefore, to efficiently identify genomic regions underlying local adaptation, a combination of genome-wide scans and GWAS can be used. In **Chapter 1** of my thesis, I identified several candidate transcripts under local adaptation at the regulatory and sequence level using a genome-wide scan approach. Three candidate transcripts that I identified to be under local adaptation at the regulatory level (H.value ≥ 0.95) overlapped with the **Chapter 3** GWAS univariate analysis results (one in control environment, two in fish environment). These three transcripts (a32_f_282383, t25908c0t3, a40_f_658599) were also under balancing selection as inferred from LOSITAN analysis. Similarly, candidates identified to have a transcriptomic basis of phenotypic variation in the GWAS multivariate analysis overlapped with 23 transcripts that I identified to have a transcriptomic basis of local adaptation at the regulatory level (H.value ≥ 0.95). Among the 23 transcripts, five transcripts were under balancing selection and two transcripts were under diversifying selection as inferred from LOSITAN analysis. These results have important implications for understanding the transcriptomic basis of local adaptation and phenotypic variation. The two transcripts that I identified in **Chapter 1** under diversifying selection and **Chapter 3** GWAS multivariate analysis implies that these transcripts lead to an increase in genetic variance in the populations as seen for example in Darwin's finches (Grant & Grant 2006). On the other hand, the eight transcripts that were under balancing selection and overlapped with the analyses of **Chapter 1** and **Chapter 3** of my thesis, might

lead to a decrease in transcriptomic variation and an average phenotype might be favored. However, further analysis and data (e.g. over time and space) are required to pinpoint the roles of different forms of selection affecting these genotype and phenotypic data.

D.3 Functional annotation

The functional impact of the vast number of genes/transcripts in *Daphnia* remains unknown, which represents a critical knowledge gap for understanding the ecological roles of candidate genes of interest. To overcome this lack of functional annotation of genes in *Daphnia*, I developed the *Daphnia* stressor database (**Chapter 2**). This database has improved our understanding of the ecological impact of *Daphnia*. For example, in **Chapter 1** of my thesis, I annotate the candidate transcripts of interest using various functional tools. However, after using the database I present in **Chapter 2**, the annotation of several "unknown" transcripts were improved, and these are evident in the annotation results I present in **Chapter 3**. Thus, even increasing our ability to know the roleplay by a fraction of a percentage is important for organisms like *Daphnia*. There are several advantages of building a functional repository exclusively for *Daphnia* based on the gene expression patterns (i) it is the first of its kind for *Daphnia* species that deals exclusively with gene expression data; (ii) the information available in the database can be used to compare to not only *Daphnia*, but also be extended to other crustaceans.

However, further work is required to enhance the information provided in the database. For example, the database in its current form is suitable only when the gene IDs or stressors or interesting literature is known. By including the function of the genes i.e., adding an extra information field with details such as "trypsin coding gene" or "oogenesis", etc., we can improve the visibility and usage statistics of the database.

Future directions

I revealed several candidate transcripts potentially under local adaptation in European *D. galeata* populations at the regulatory and sequence level (**Chapter 1**). These results are an important resource for future research on local adaptation in *Daphnia* populations. The identified candidate transcripts for local adaptation can be used to test and compare against any *Daphnia* stressors using comparative genomics approaches. Genome-wide data should be made available to enable comparisons with the transcriptome data used in the present study, ideally across multiple stressors. This information would help to assess the combination of both regulatory and sequence level variation and their contributions to population divergence across multiple stressors. Genome-wide data can also enhance our understanding of other events such as gene duplication, splicing and the role of non-coding regions, which can aid in assessing local adaptation and genotype-phenotype correlations more precisely. The candidate transcripts that I identified to have a genetic basis of phenotypic variation (**Chapter 3**) can be used as candidates for understanding the interactions between prey and predator in aquatic environments. Furthermore, the candidate transcripts I identified in **Chapter 1** and **Chapter 3** can act as the basis for testing other ecologically relevant stressors not only in *D. galeata* but also in other *Daphnia* species and insects, as information on the orthologs and paralogs are readily available. Finally, the *Daphnia* stressor database I developed (**Chapter 2**) can be used to shortlist genes of interest based on stressors and use them as candidates in further research and test functionally important genes in *Daphnia*. This current database should be extended as and when new gene expression data is published and keep up to date information on the adaptive genes under various environmental conditions. Furthermore, adding functional information such as gene ontology or protein domain annotations can drastically enhance the usage of the database.

Conclusions

The utilization of ecologically relevant genes in *Daphnia* might reveal general patterns of adaptive processes and evolutionary divergence, which are important aspects of understanding biodiversity and habitat fragmentation. For example, gene flow between populations and species reduces the risk of local extinction but might alter the genetic makeup of populations. This acts as a basis for selection and adaptive divergence when under constant environmental disturbances. However, genetic evidence for habitat fragmentation driven evolution is still sparse and there are two general explanations by evolutionary biologists as to why this is the case. First, the methodological and technical difficulties along with a lack of long-term data to understand the history of organisms and habitats are essential to predict the past and future of habitat fragmentation driven evolution. Secondly, biological phenomena such as genetic constraints and ecological niche conservatism might be another plausible and mutually non-exclusive explanation (Merila 2012) as to why genetic evidence is lacking. However, although phenotypic plasticity along with dispersal might provide favorable conditions and act as a buffer against fitness losses due to changing environmental conditions, they are unlikely to provide long term species sustainability unless plasticity is the major factor influencing evolutionary adaptation through genetic assimilation (Valladares *et al*. 2014). Although several studies have provided insights into habitat fragmentation and evolutionary adaptation, it's still far from making general predictions about how populations adapt to their environments and persist in the long run. In this view, the words from Holt made in 1990, still holds true today "There is almost no species for which we know enough relevant ecology, physiology and genetics to predict its evolutionary response to climate change". However, with the recent advancements in high-throughput sequencing technologies and research on several organisms to understand genetic and phenotypic variation, I am positive that we will improve our predictions and hypothesis that we put forward to assess the role of ecology, physiology and genetics to predict evolutionary changes.

Appendix for Introduction

Image sources for Figure 2:

[1]: http://canopusglobal.com/services/agrochemicals/
[2]: http://www.chemspider.com/Chemical-Structure.13837988.html
[3]: https://www.thoughtco.com/list-of-common-polyatomic-ions-603977
[4]: http://csabusiness.com/bangladesh-high-court-bans-20-pharmaceutical-companies
[5]: https://blog.bestbuy.ca/appliances/major-appliances/what-temperature-should-your-fridge-and-freezer-be
[6]: https://de.wikipedia.org/wiki/Pasteuria_ramosa#/media/File:Pasteuria_ramosa_spores.jpg
[7]: http://www.marinephytoplankton.org/tag/marine-phytoplankton-benefits/
[8]: https://nature.mdc.mo.gov/discover-nature/field-guide/bluegill
[9]: https://en.wikipedia.org/wiki/Glassworm#/media/File:GlasswormLateralView.JPG
[10]: https://en.wikipedia.org/wiki/Zooplankton#/media/File:Zooplankton.jpg

Appendix for Chapter 1

Figure A1_F1: Gene expression PCA for the first three principal components. Gene expression PCA of the four sampled populations for the first three principal components pairwise as a matrix; Pop.G (Lake Greifensee), Pop.J (Jordan Reservoir), Pop.LC (Lake Constance) and Pop.M (Lake Müggelsee). Percentages on the X- and Y-axis indicate the percentage of variance explained by each principal component.

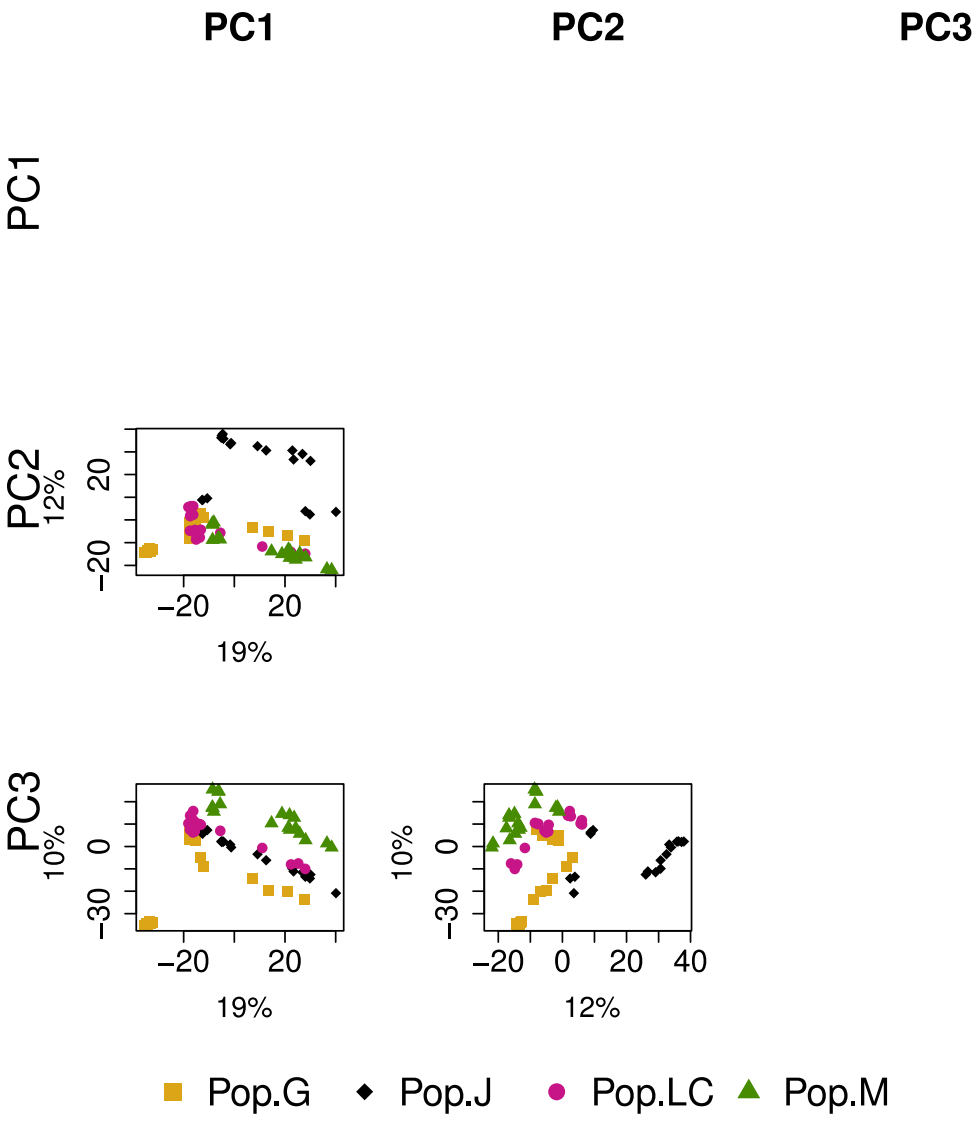


Figure A1_F2: SNP PCA for the first three principal components. SNP PCA of the four sampled populations for the first three principal components pairwise as a matrix; Pop.G (Lake Greifensee), Pop.J (Jordan Reservoir), Pop.LC (Lake Constance) and Pop.M (Lake Müggelsee). Percentages on the X- and Y-axis indicate the percentage of variance explained by each principal component.

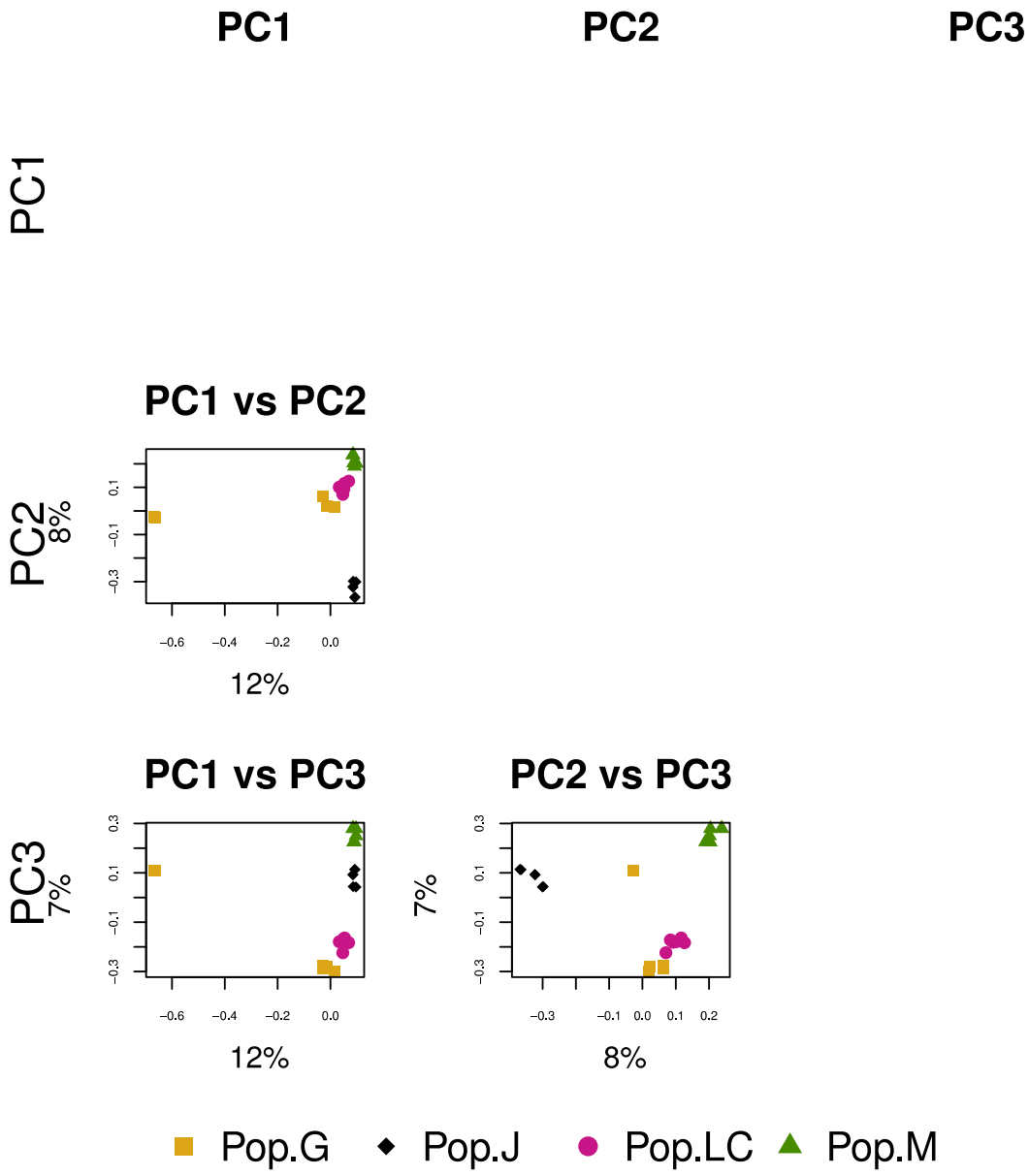
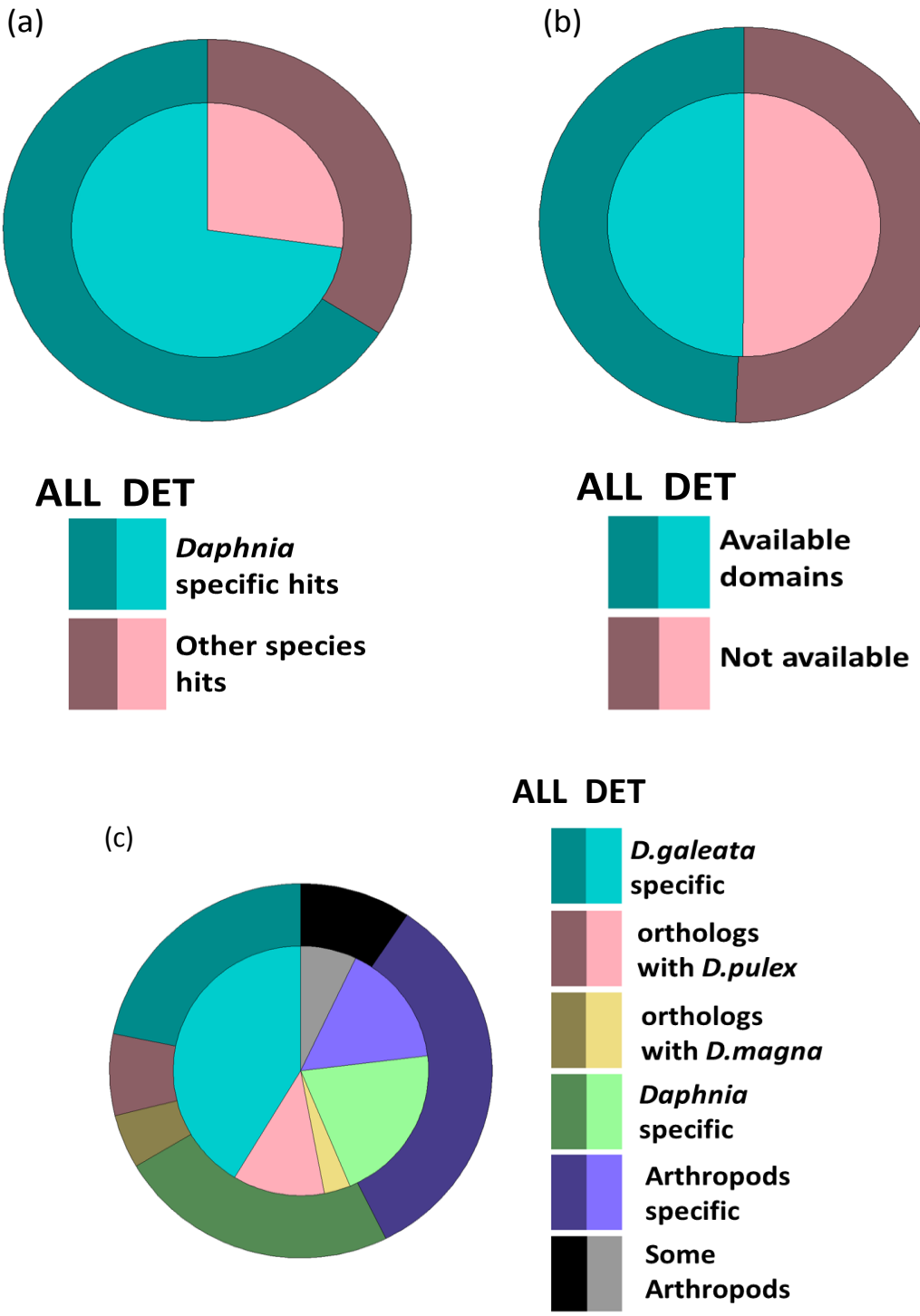


Figure A1_F3a-c: Pie charts showing the number of transcripts annotated using BLAST, Pfam and orthoMCL analysis. The outer circle represents the data for all 32903 transcripts. The inner circle represents the data for the differentially expressed transcripts. (a) Pie chart for (a) BLAST; (b) Pfam (c) orthoMCL



Appendix for Chapter 3

Table A3_S3: GWAS results of "inflated" dataset in control, fish environments and gxe interaction.

	Control: $-\log_{10}P \geq 1.5$	Fish: $-\log_{10}P \geq 1.5$	gxe: $-\log_{10}P \geq 1.5$
brood1_snps	5258	7309	0
brood1_transcripts	2457	3049	0
brood2_snps	13018	14420	8
brood2_transcripts	4686	5008	5
brood3_snps	11231	18383	49
brood3_transcripts	4206	5716	40
brood4_snps	3	7	0
brood4_transcripts	3	7	0
afr_snps	165	8625	24
afr_transcripts	115	3071	20
broods_snps	2656	3306	74
broods_transcripts	1414	1652	50
offspring_snps	23284	32126	0
offspring_transcripts	6374	7486	0
survival_snps	0	0	6
survival_transcripts	0	0	3
length_snps	22085	19998	23
length_transcripts	6335	6026	21
sgr_snps	18763	18838	10
sgr_transcripts	5774	5837	10

Table A3_S5: Functional annotation of candidate transcripts of interest as identified in the univariate and multivariate GWAS analysis and WGCNA. A total of 156 candidate transcripts are listed. 'orthogroup' refers to the orthoMCL cluster the transcript is assigned to, 'dgal' = number of *D. galeata* transcripts present in the assigned orthoMCL cluster. 'dpul' = number of *D. pulex* genes present in the assigned orthoMCL cluster. 'dmag' = number of *D. magna* genes present in the assigned orthoMCL cluster. 'dme' = number of *Drosophila melanogaster* genes present in the assigned orthoMCL cluster. 'nvi' = number of *Nasonia vitripennis* genes present in the assigned orthoMCL cluster. 'hit' refers to a transcript that is significantly homologous in *Daphnia* stressor database to the candidate transcript of interest. 'identity_percent' refers to BLAST identity percentage corresponding to the *Daphnia* stressor database hit.

Transcripts	Analysis	Environment	Trait	orthogroup	dgal	dpul	dmag	dme	nvi	GO_annotation	tBLASTx			BLASTx		
											tBLASTx_hit	tBLASTx_identity_percentage	Stressor_tBLASTx	BLASTx_hit	BLASTx_identity_percentage	Stressor_BLASTx
t25908c0t3	GWAS_univariate	fish	brood4	ORTHO_ALL527	2	1	1	0	1							
s324748	GWAS_univariate	control	brood4	ORTHO_ALL18855	1	0	0	0	0							
a32_f_282383	GWAS_univariate	control	brood4	ORTHO_ALL12465	1	1	0	0	0	GO:0004222, GO:0006508,	DappuDraft_302051	59.21	Salinity			
a22_f_855011	GWAS_univariate	fish	brood3	ORTHO_ALL268	6	4	4	4	5	GO:0005524, GO:0016887,						
a40_f_658599	GWAS_univariate	fish	brood3	ORTHO_ALL5300	1	1	1	1	1							
t25796c0t2	GWAS_univariate	fish	brood3	ORTHO_ALL6438	1	1	1	1	1	GO:0005515,						
t15632c0t6	GWAS_univariate	gxe	offspring	ORTHO_ALL5309	1	1	2	0	0	GO:0007283,						
o2202d27113t4	GWAS_multivariate	control	all	ORTHO_ALL7667	1	1	1	0	0							
s398707	GWAS_multivariate	control	all	ORTHO_ALL344	18	0	0	0	0							
t24633c0t1	GWAS_multivariate	control	all	ORTHO_ALL588	2	3	3	0	0	GO:0004553, GO:0005975,	DappuDraft_326098	72.19	Light Dark Cycle			
o6273d6138t5	GWAS_multivariate	control	all	ORTHO_ALL590	2	2	2	2	1	GO:0016491, GO:0055114,	DappuDraft_188248	70.96	Phosphorous			
o1345t5	GWAS_multivariate	control	all	ORTHO_ALL301	2	4	4	3	4	GO:0003774, GO:0005524, GO:0016459, GO:0005515, GO:0005856,						
s378541	GWAS_multivariate	control	all	ORTHO_ALL10824	1	1	1	0	0	GO:0003777, GO:0005524, GO:0008017, GO:0007018,						
a84_f_13467	GWAS_multivariate	control	all	ORTHO_ALL8014	1	2	2	0	0							
s449363	GWAS_multivariate	control	all	ORTHO_ALL542	3	3	5	1	1	GO:0003824, GO:0008152,	Dapma7bEVm004083t1	67.48	Tri Nitro Toluene			
o1384d26522t2	GWAS_multivariate	control	all	ORTHO_ALL5612	1	1	1	1	1	GO:0005515, GO:0005737,						
t25844c0t12	GWAS_multivariate	control	all	ORTHO_ALL333	4	2	3	3	2	GO:0005509, GO:0007156, GO:0016020, GO:0005515,						
o5466t2	GWAS_multivariate	control	all	ORTHO_ALL3010	1	2	1	1	1							
t20144c2t5	GWAS_multivariate	control	all	ORTHO_ALL4658	1	2	1	1	1	GO:0005515,						
t25497c0t6	GWAS_multivariate	control	all	ORTHO_ALL390	1	1	1	1	0							
o4766d10624t1	GWAS_multivariate	control	all	ORTHO_ALL7635	1	1	0	0	0							
o5937t7	GWAS_multivariate	control	all	ORTHO_ALL342	5	7	7	3	2	GO:0005328, GO:0006836, GO:0016021,						
t22913c0t7	GWAS_multivariate	control	all	ORTHO_ALL6660	1	1	1	2	1	GO:0004332, GO:0006096,						
t20791c0t3	GWAS_multivariate	control	all	ORTHO_ALL1143	2	2	1	1	2	GO:0004672, GO:0005524, GO:0006468,						
t24368c0t5	GWAS_multivariate	control	all	ORTHO_ALL152	7	1	2	0	0	GO:0003676,						
o9097d30577t2	GWAS_multivariate	control	all	ORTHO_ALL838	1	4	2	0	0	GO:0005515, GO:0000228,						
t24099c0t1	GWAS_multivariate	control	all	ORTHO_ALL3185	2	1	1	0	0							
o11620d39461t2	GWAS_multivariate	control	all	ORTHO_ALL5376	1	1	1	1	1	GO:0005515,						
t24918c0t1	GWAS_multivariate	control	all	ORTHO_ALL367	4	5	3	0	1	GO:0055085, GO:0016021,						
t25138c1t20	GWAS_multivariate	control	all	ORTHO_ALL4665	1	1	1	0	0							
t21653c0t1	GWAS_multivariate	control	all	ORTHO_ALL6333	1	1	1	1	1							
a34_j_729597	GWAS_multivariate	gxe	all	ORTHO_ALL2541	1	1	1	1	2	GO:0003723, GO:0006396, GO:0005515,						
t20056c0t3	GWAS_multivariate	gxe	all	ORTHO_ALL7934	1	1	0	1	1	GO:0004222, GO:0006508,	DappuDraft_254737	74.11	Salinity; Light Dark Cycle; Phosphorous			
a34_f_231353	GWAS_multivariate	gxe	all	ORTHO_ALL10125	1	1	1	0	0							
s324748	GWAS_multivariate	gxe	all	ORTHO_ALL18855	1	0	0	0	0							
a60_f_293849	GWAS_multivariate	gxe	all	ORTHO_ALL15855	1	0	0	0	0	GO:0004672, GO:0006468,						

t22396c0t2	GWAS_multivariate	gxe	all	ORTHO_ALL3492	1	2	3	0	0							
s390254	GWAS_multivariate	gxe	all	ORTHO_ALL14997	1	1	1	0	0							
t23826c0t5	GWAS_multivariate	gxe	all	ORTHO_ALL10752	1	1	1	0	0							
t23664c0t1	GWAS_multivariate	gxe	all	ORTHO_ALL7919	1	1	1	0	0							
o7642t5	GWAS_multivariate	gxe	all	ORTHO_ALL7034	3	0	0	0	0							
a32_f_282383	GWAS_multivariate	gxe	all	ORTHO_ALL12465	1	1	0	0	0	GO:0004222, GO:0006508,	DappuDraft_302051	59.21	Salinity			
t14146c0t1	GWAS_multivariate	gxe	all	ORTHO_ALL11216	1	0	3	0	0	GO:0008417, GO:0006486, GO:0016020,						
sd58198123	GWAS_multivariate	gxe	all	ORTHO_ALL2635	2	1	1	1	1	GO:0005515,						
a26_f_347901	GWAS_multivariate	gxe	all	ORTHO_ALL581	5	6	6	1	1	GO:0005506, GO:0016705, GO:0020037, GO:0055114,						
t19638c1t2	GWAS_multivariate	gxe	all	ORTHO_ALL3311	1	1	1	1	1	GO:0004114, GO:0007165,	DappuDraft_321218	63.61	Phosphorous			
t19154c0t4	GWAS_multivariate	gxe	all	ORTHO_ALL3403	1	1	1	1	1	GO:0002161,						
t18105c1t3	GWAS_multivariate	gxe	all	ORTHO_ALL3317	1	2	2	1	1	GO:0008270,						
o5648d38044t1	GWAS_multivariate	gxe	all	ORTHO_ALL10025	1	1	1	0	0							
t25497c0t6	GWAS_multivariate	gxe	all	ORTHO_ALL390	1	1	1	1	0							
a9699	GWAS_multivariate	gxe	all	ORTHO_ALL5025	1	1	1	1	1	GO:0005525, GO:0005634, GO:0005737,						
o16998d40345t1	GWAS_multivariate	gxe	all	ORTHO_ALL4372	1	0	0	0	0							
o11352t2	GWAS_multivariate	gxe	all	ORTHO_ALL14353	1	1	1	0	0	GO:0004797, GO:0005524,						
t11417c0t1	GWAS_multivariate	gxe	all	ORTHO_ALL15047	1	1	1	0	0							
o7501d29931t2	GWAS_multivariate	gxe	all	ORTHO_ALL510	1	1	0	0	0	GO:0004222, GO:0008270, GO:0006508, GO:0031012,						
a3928	GWAS_multivariate	gxe	all	ORTHO_ALL7076	1	1	1	0	0	GO:0005515,	DappuDraft_319446	56.15	Temperature			
a24_f_775461	GWAS_multivariate	gxe	all	ORTHO_ALL5191	1	1	1	1	1							
s392443	GWAS_multivariate	gxe	all	ORTHO_ALL12802	1	1	0	0	0							
t24609c0t3	GWAS_multivariate	gxe	all	ORTHO_ALL2402	1	1	1	1	2		Dapma7bEVm005424t1	64.94	Tri Nitro Toluene			
o5579t1	GWAS_multivariate	gxe	all	ORTHO_ALL918	4	3	2	1	1	GO:0005515, GO:0004435, GO:0006629, GO:0007165, GO:0035556,						
s389937	GWAS_multivariate	gxe	all	ORTHO_ALL1425	1	1	1	1	1	GO:0016020,	DappuDraft_191157	52.78	Phosphorous			
t12328c0t2	GWAS_multivariate	gxe	all	ORTHO_ALL10265	1	1	1	0	0	GO:0003700, GO:0043565, GO:0006355,						
o5887d17451t1	GWAS_multivariate	gxe	all	ORTHO_ALL1702	1	2	1	1	1	GO:0005515, GO:0004672, GO:0006468,						
o13048d20975t1	GWAS_multivariate	gxe	all	ORTHO_ALL8804	1	1	1	1	1							
o5843d43690t1	GWAS_multivariate	gxe	all	ORTHO_ALL253	3	4	6	4	3	GO:0035556, GO:0005515, GO:0004672, GO:0005524, GO:0006468,						
t21598c0t2	GWAS_multivariate	gxe	all	ORTHO_ALL24	20	6	22	0	0	GO:0005515,						
o1345t5	GWAS_multivariate	gxe	all	ORTHO_ALL301	2	4	4	3	4	GO:0003774, GO:0005524, GO:0016459, GO:0005515, GO:0005856,						
t18365c0t1	GWAS_multivariate	gxe	all	ORTHO_ALL6050	1	2	1	0	1	GO:0004866, GO:0005615, GO:0005576,						
t25696c1t2	GWAS_multivariate	gxe	all	ORTHO_ALL2696	1	1	3	1	1	GO:0003677, GO:0006355,						
t24931c0t1	GWAS_multivariate	gxe	all	ORTHO_ALL1336	1	1	1	1	1	GO:0045454,						
t25838c2t1	GWAS_multivariate	gxe	all	ORTHO_ALL2522	1	1	4	1	1	GO:0005509, GO:0005515,						
sd44175363445	WGCNA	control	broods	ORTHO_ALL156	15	0	0	0	0	GO:0004252, GO:0006508,						
t17566c0t1	WGCNA	control	broods	ORTHO_ALL10786	1	1	1	0	0							
s312581	WGCNA	control	broods	ORTHO_ALL280	11	0	0	0	0							

o21110d23720t1	WGCNA	control	broods							GO:0008277,						
s417471	WGCNA	control	broods	ORTHO_ALL280	11	0	0	0	0							
a84_f_236803	WGCNA	control	broods	ORTHO_ALL187	20	0	0	0	0							
t21568c0t2	WGCNA	control	broods													
t42914c0t1	WGCNA	control	broods													
a36_b_702278	WGCNA	control	broods	ORTHO_ALL404	7	0	0	0	0	GO:0004672, GO:0005524, GO:0006468,						
s391499	WGCNA	control	broods	ORTHO_ALL15593	1	0	0	0	0		Dapma7bEVm000945t1	62.20	Chlamydomonas; Microcystin; Lead; Tri Nitro Toluene			
sd30928308739	WGCNA	control	broods	ORTHO_ALL166	21	1	0	0	0	GO:0004672, GO:0005524, GO:0006468,						
t22220c0t1	WGCNA	control	broods	ORTHO_ALL125	15	3	2	0	0							
a60_j_540606	WGCNA	control	broods	ORTHO_ALL2147	1	1	2	1	1		Dapma7bEVm015471t1	50.59	Acrylonitrile; Chloro Vinyl Ether; 20-hydroxyecdysone; Phenol; MTBE; PonasteroneA; Trichloroethylene; Atrazine; Dichlorobenzene; Beta Estradiol; Parathion; Diazinon; Phenanthrene; Pyripoxyfen; Methoxychlor; Toxaphene; Bifenthrin; Lambda Cyhalothrin; Permethrin; Tri Nitro Toluene;			
t21412c0t10	WGCNA	control	broods	ORTHO_ALL1050	3	1	2	5	0	GO:0004222, GO:0006508,	DappuDraft_93694	62.73	Salinity	DAPPUDRAFT_347623	63.75	hypoxia
t17214c0t3	WGCNA	control	broods	ORTHO_ALL940	2	1	0	0	0							
o18816t1	WGCNA	control	broods	ORTHO_ALL19804	1	0	1	0	0	GO:0004550, GO:0005524, GO:0006165, GO:0006183, GO:0006228, GO:0006241,	NM_008704.2	62.02	Dinitrobenzene			
a22_f_12916	WGCNA	control	broods													
t30564c0t1	WGCNA	control	broods	ORTHO_ALL8054	1	0	4	0	0							
t37310c0t1	WGCNA	control	broods													
a5310	WGCNA	control	broods	ORTHO_ALL737	9	1	0	0	0	GO:0005515,						
o851t2	WGCNA	control	broods	ORTHO_ALL9371	1	1	1	1	1	GO:0005524, GO:0042626, GO:0006810, GO:0055085, GO:0016021, GO:0016887,	Dapma7bEVm001092t1	58.78	Tri Nitro Toluene			
a7590	WGCNA	control	broods	ORTHO_ALL12363	2	0	0	0	0							
s359419	WGCNA	control	broods	ORTHO_ALL4987	2	0	0	0	0	GO:0004672, GO:0006468,						
t20470c0t1	WGCNA	control	broods	ORTHO_ALL13117	2	0	0	0	0							
sd39683267849	WGCNA	control	broods													
t49177c0t1	WGCNA	control	broods													
t34171c0t1	WGCNA	control	broods													
a40_f_333407	WGCNA	control	broods	ORTHO_ALL779	8	1	2	0	0							
t36041c0t1	WGCNA	control	broods	ORTHO_ALL9848	1	1	1	1	1	GO:0003700, GO:0043565, GO:0006355,						
o2379t3	WGCNA	control	broods	ORTHO_ALL3762	1	1	1	1	2	GO:0015116, GO:0008272, GO:0016021,						

o2330t1	WGCNA	control	broods	ORTHO_ALL2424	1	1	1	1	1	GO:0006508,						
o9767d892t2	WGCNA	control	broods													
t23036c0t1	WGCNA	control	broods	ORTHO_ALL4199	1	1	1	1	1							
a5709	WGCNA	control	broods	ORTHO_ALL4983	2	0	0	0	0	GO:0003676, GO:0005524,						
a5632	WGCNA	control	broods													
s233253	WGCNA	control	broods											ABD19215	73.93	Cadmium
a84_f_221570	WGCNA	control	broods	ORTHO_ALL7	124	0	0	0	0		DappuDraft_257184	52.94	Temperature			
a76_f_462372	WGCNA	control	broods	ORTHO_ALL5561	1	1	1	1	1							
o395t1	WGCNA	control	broods							GO:0008483, GO:0030170,	Dapma7bEVm010837t1	60.36	Acrylonitrile; Chloro Vinyl Ether; 20-hydroxyecdysone; Phenol; MTBE; Chloroform; PonasteroneA; Trichloroethylene; Atrazine; Dichlorobenzene; Beta Estradiol; Diazinon; Phenanthrene; Methoxychlor; Chlorpyrifos; Toxaphene; Methyl Farnesoate; Bifenthrin; Lambda Cyhalothrin; Nonylphenol; Permethrin; Tri Nitro Toluene; Epofenonane;			
t24577c0t1	WGCNA	control	broods	ORTHO_ALL2100	1	2	0	0	0							
t20212c0t4	WGCNA	control	broods	ORTHO_ALL6418	5	0	0	0	0							
t33183c0t1	WGCNA	control	broods													
o6289d17688t1	WGCNA	control	broods	ORTHO_ALL19907	1	1	0	0	0	GO:0005515,						
sd45262380885	WGCNA	control	broods													
sd23587267991	WGCNA	control	broods													
s371088	WGCNA	control	broods	ORTHO_ALL1350	3	1	2	0	0	GO:0004930, GO:0007186, GO:0016021,						
t34628c0t1	WGCNA	control	broods													
sd557203853	WGCNA	control	broods	ORTHO_ALL7014	1	2	1	1	1	GO:0003677,						
sd14317394433	WGCNA	control	broods													
o2927d7883t2	WGCNA	control	broods	ORTHO_ALL48	27	0	0	0	0							
s375229	WGCNA	control	broods	ORTHO_ALL398	9	0	0	0	0							
o3698t3	WGCNA	control	broods	ORTHO_ALL324	10	3	2	0	1		hxAUG25s183g211t1	67.59	Fish kairomones			
s378521	WGCNA	control	broods	ORTHO_ALL5	87	5	5	0	0							
o17558t2	WGCNA	control	broods	ORTHO_ALL8912	1	1	1	1	1							
o3505t1	WGCNA	control	broods	ORTHO_ALL38	38	0	0	0	0							
o1704d1887t1	WGCNA	control	broods	ORTHO_ALL3008	1	1	2	1	1	GO:0019901, GO:0000079,						
o2936t1	WGCNA	control	broods	ORTHO_ALL5	87	5	5	0	0							
o13262t1	WGCNA	control	broods	ORTHO_ALL3055	1	1	2	0	0	GO:0008483, GO:0030170,	DappuDraft_209533	67.61	Phosphorous			
s315181	WGCNA	control	broods	ORTHO_ALL3	138	3	1	1	0	GO:0004672, GO:0006468,						

t25809c0t15	WGCNA	control	broods	ORTHO_ALL13807	2	0	0	0	0							
a7623	WGCNA	control	broods	ORTHO_ALL166	21	1	0	0	0	GO:0004672, GO:0005524, GO:0006468,						
t32557c0t1	WGCNA	control	broods													
s309681	WGCNA	control	broods													
s422203	WGCNA	control	broods	ORTHO_ALL24	20	6	22	0	0	GO:0005515,						
sd32516319717	WGCNA	control	broods	ORTHO_ALL9	103	2	0	0	0							
s445117	WGCNA	control	broods	ORTHO_ALL459	12	0	0	0	0							
o11854t1	WGCNA	control	broods	ORTHO_ALL63	17	18	19	0	1	GO:0004672, GO:0006468,						
s297019	WGCNA	control	broods	ORTHO_ALL133	20	0	0	0	0							
t30618c0t1	WGCNA	control	broods													
a22_f_206059	WGCNA	control	broods	ORTHO_ALL18226	1	0	0	0	0							
a34_b_719141	WGCNA	control	broods													
s363545	WGCNA	control	broods	ORTHO_ALL13080	1	0	0	0	0							
s236301	WGCNA	control	broods													
t5465c0t1	WGCNA	control	broods													
t37736c0t1	WGCNA	control	broods													
s337106	WGCNA	control	broods													
s356485	WGCNA	control	broods													
t25768c0t5	WGCNA	control	broods	ORTHO_ALL63	17	18	19	0	1							
s340859	WGCNA	control	broods	ORTHO_ALL15628	1	0	0	0	0							
t47565c0t1	WGCNA	control	broods	ORTHO_ALL2816	1	7	1	0	0		DappuDraft_328985	76.52	Temperature			
s323057	WGCNA	control	broods	ORTHO_ALL207	16	4	0	0	0							

Figure A3_S1: Breeding design of life-history experiment in the absence or presence of fish kairomones (Tams *et al* submitted). Each clonal line was bred in kairomone-free water (control) and in kairomone water (fish) for two subsequent generations (F0 & F1). Neonates from 3rd to 5th brood were used to start a new generation. Life-history traits of experimental individuals (F2) were measured for 14 days (t14). Neonates were preserved in ethanol at the beginning of the experiment (t0) and experimental individuals on day 14 (t14) to measure the trait size (body length).

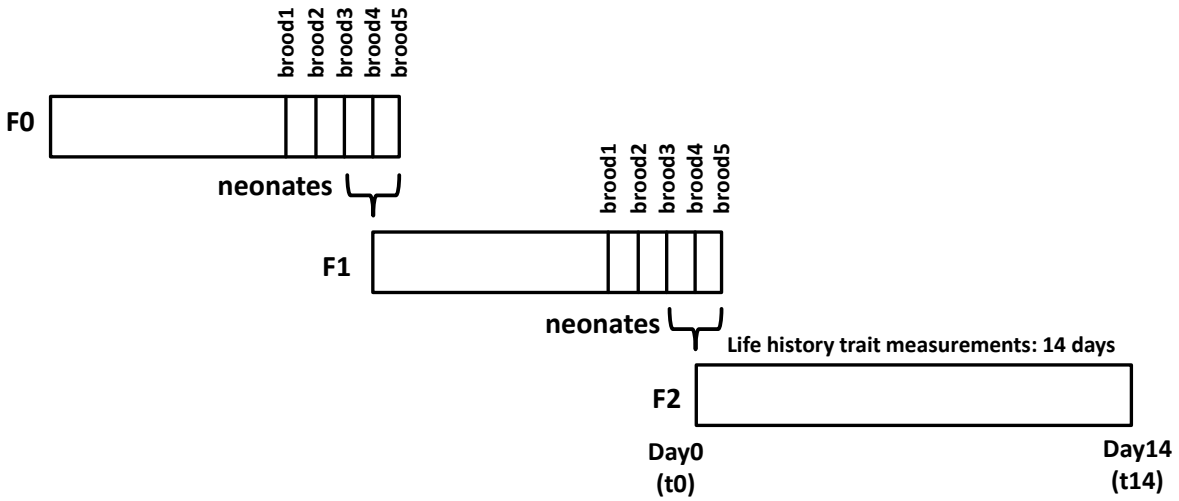


Figure A3_S2: A visual representation of how the SNP dataset was created for "inflated" GWAS analysis.

SNP	Ref	Alt	PopG						PopJ						PopLC						PopM					
			G1.11	G1.12	G1.6	G1.7	G2.1	G3.1	J1	J2	J2.1	J2.4	J3	J4	LC3.1	LC3.3	LC3.5	LC3.6	LC3.7	LC3.9	M10	M12	M2	M5	M6	M9
abyssk22_f_5797:149	G	C	0/1	0/1	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0
trinitytrinloc24474c0t10:96	A	G	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/1
oasetrinityLoc4568d28579t2:4039	T	G	0/0	0/0	0/1	0/1	0/0	0/2	0/2	0/0	0/0	0/0	0/2	0/0	0/0	0/2	0/0	0/0	0/2	0/0	0/2	0/2	0/2	0/0	0/2	0/2
soapsoap366791:179	A	G	0/0	0/0	0/1	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0

Control condition

	G.11																									
Individual	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt
abyssk22_f_5797:149	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C

Fish kairomone condition

	G.11																									
Individual	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt	Ref	Alt
abyssk22_f_5797:149	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C	G	C

Author Contributions

The work presented in **Chapter-1** was submitted to *Molecular Ecology* as "**Contrasting patterns of divergence at the regulatory and sequence level in European *Daphnia galeata* natural populations**" by **Suda Parimala Ravindran** (author of this declaration), Maike Herrmann and Mathilde Cordellier (thesis supervisor). This study was conceived by Mathilde Cordellier and me. I performed all the statistical data analysis except for LOSITAN and wrote the manuscript.

The work presented in **Chapter-2** will be submitted to *Nature Scientific Reports* as "***Daphnia* stressor database: Taking advantage of a decade of *Daphnia* '-omics' data for gene annotation**" by **Suda Parimala Ravindran** (author of this declaration), Lisa Gottschlich, Jennifer Lüneburg, Verena Tams and Mathilde Cordellier (thesis supervisor). This study was designed by Mathilde Cordellier and me. Literature for the meta-analysis was collected by Mathilde Cordellier and me. All authors were involved in data mining. I compiled, analyzed the data, developed the website and wrote the manuscript.

The work presented in **Chapter-3** will be submitted for publication in Genes, Genomes and Genetics as "**An environment-dependent genotype-phenotype association in European *Daphnia galeata***" by **Suda Parimala Ravindran** (author of this declaration), Verena Tams and Mathilde Cordellier (thesis supervisor). This study was conceived by Mathilde Cordellier, Verena Tams and me. I conducted the genome-wide association analyses and functional annotation. Verena Tams, Jennifer Lüneburg and Laura Seddar performed the life-history trait experiments. Verena Tams performed the gene co-expression network analysis. I wrote the manuscript along with Verena Tams.

Eidesstattliche Versicherung

Declaration on oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den Unterschrift

Acknowledgements

First and foremost, I would like to thank my PhD supervisor **Mathilde Cordellier** for accepting me as her PhD student and for her support and guidance throughout these years! She was not just my supervisor, but also a person to whom I could always run to when I encountered work-related issues. Research seemed easier and understandable after having a word with you. Thanks for believing in me, sometimes more than I did (especially with Circos plots :-P). I would definitely consider her as one of my most favorite supervisors in my life! If at all I chose to be in academia in the future, I would look upon her as my role model. Thank you for everything, Mathilde!

Second, I would like to thank **Dr. Luisa Orsini** for readily accepting my request to be one of my thesis examiners.

Next I would like to thank my office mates, **Verena** and **Jenni** for being the official translators of all my documents from German to English. I had a lot of fun jabbering about many things about life, many (secret) gossips and for sharing the same enthusiasm when we heard the word "food". My days at work would have been much lonelier and boring without you two!

Many thanks to Prof. Susanne Dobler and her working group for being my general audience whenever I had to present my data and for their valuable suggestions to improve my talk for conferences.

Friends make my life complete. They have been my biggest source of strength and I thank them for always being one text/phone call away. Our never-ending discussions, gossips, movies, music, cricket, spiritual stuff and a lot more have always kept me going and face the typical trauma of a PhD student.

Next, I would like to thank my dad, **Ravindran** (daddy) and my mother, **Premalatha** (amma) for always providing only the best in my life and for their constant support, love, encouragement and prayers every day. Our bonding only grew fonder despite being ~8000kms away from each other!

I thank my brother **Prabu** (Anooos), my sister-in-law **Vaishnavi** (VKD) for their support and encouragement. Special thanks to my two darling nephews, **Vishruth** (guddu) and **Shashwath** (shashyyy boy) for their non-stop entertainment, bringing a smile every time I talk (play) with them. I forget all my hardships while talking/watching them.

Last but not the least, I would like to thank the almighty **God** for showering His blessings, providing me with a lovely family, helping me cope with life and answering my prayers.

References

Abrams PA, Rueffler C, Kim G (2008) Determinants of the strength of disruptive and/or divergent selection arising from resource competition. *Evolution* **62**, 1571-1586.

Agrawal AA (2001) Phenotypic plasticity in the interactions and evolution of species. *Science* **294**, 321-326.

Alberto FJ, Aitken SN, Alia R*, et al.* (2013) Potential for evolutionary responses to climate change - evidence from tree populations. *Glob Chang Biol* **19**, 1645-1661.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5252-5257.

Alexa A, Rahnenfuhrer J (2016) topGO: Enrichment analysis for gene ontology.

Allen MR, Thum RA, Caceres CE (2010) Does local adaptation to resources explain genetic differentiation among *Daphnia* populations? *Mol Ecol* **19**, 3076-3087.

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* **5**, 181-190.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Ambrosone CB (2007) The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer. *Breast Cancer Res* **9**, 114.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* **9**, 323.

Atanasov KE, Barboza-Barquero L, Tiburcio AF, Alcazar R (2016) Genome wide association mapping for the tolerance to the polyamine oxidase inhibitor guazatine in *Arabidopsis thaliana*. *Front Plant Sci* **7**, 401.

Ayrinhac A, V. D, P. G*, et al.* (2004) Cold adaptation in geographical populations of *Drosophila melanogaster:* phenotypic plasticity is more important than genetic variability. *Functional Ecology* **18**, 700-706.

Barata C, Baird DJ, Mitchell SE, Soares AMVM (2002) Among- and within-population variability in tolerance to cadmium stress in natural populations of *Daphnia magna*. *Environmental Toxicology and Chemistry* **21**, 1058-1064.

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38-44.

Barrett T, Wilhite SE, Ledoux P*, et al.* (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* **41**, D991-995.

Beaugrand G, Kirby RR (2018) How do marine pelagic species respond to climate change? Theories and observations. *Ann Rev Mar Sci* **10**, 169-197.

Beckerman AP, Rodgers GM, Dennis SR (2010) The reaction norm of size and age at maturity under multiple predator risk. *Journal of Animal Ecology* **79**, 1069-1076.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecol Lett* **15**, 365-377.

Benjamini YH, Yosef (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* **57**, 289-300.

Bento G, Routtu J, Fields PD*, et al.* (2017) The genetic basis of resistance and matching-allele interactions of a host-parasite system: The *Daphnia magna-Pasteuria ramosa* model. *PLoS Genetics* **13**, e1006596.

Berg L (2001) *Genetic disequilibria and the interpretation of population genetic structure in Daphnia*, Uppsala
Black AR, Dodson SI (1990) Demographic costs of *Chaoborus*-induced phenotypic plasticity in *Daphnia pulex*. *Oecologia* **83**, 117-122.
Boake CRB, Gavrilets S (1998) On the evolution of premating isolation after a founder event. *The American Naturalist* **152**, 706-716.
Boeing WJ, Ramcharan CW, Riessen HP (2006) Multiple predator defence strategies in *Daphnia pulex* and their relation to native habitat. *Journal of Plankton Research* **28**, 571-584.
Boersma M, Spaak P, De Meester L (1998) Predator-mediated plasticity in morphology, life history, and behavior of *Daphnia*: the uncoupling of responses. *The American Naturalist* **152**, 237-248.
Boileau MG, Hebert PDN, Schwartz SS (1992) Non-equilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology* **5**, 25-39.
Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics* **38**, 459-487.
Böndel KB, Lainer H, Nosenko T*, et al.* (2015) North-South colonization associated with local adaptation of the wild tomato species *Solanum chilense*. *Molecular Biology and Evolution* **32**, 2932-2943.
Bonte D, Dahirel M (2017) Dispersal: a central and independent trait in life history. *Oikos* **126**, 472-479.
Botkin DB, Saxe H, Araújo MB*, et al.* (2007a) Forecasting the effects of global warming on biodiversity. *BioScience*.
Botkin DB, Saxe H, Araújo MB*, et al.* (2007b) Forecasting the effects of global warming on biodiversity. *BioScience* **57**, 227-236.
Bourguet D (1999) The evolution of dominance. *Heredity* **83**.
Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews* **80**, 205-225.
Bowman LL, Kondrateva ES, Timofeyev MA, Yampolsky LY (2018) Temperature gradient affects differentiation of gene expression and SNP allele frequencies in the dominant Lake Baikal zooplankton species. *Mol Ecol* **27**, 2544-2559.
Brede N, Sandrock C, Straile D*, et al.* (2009) The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4758-4763.
Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1572-1577.
Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752-755.
Brennan RS, Healy TM, Bryant HJ*, et al.* (2018) Genome-wide selection scans integrated with association mapping reveal mechanisms of physiological adaptation across a salinity gradient in killifish. *biorxiv*.
Brown WLJ, Eisner T, Whittaker RH (1970) Allomones and kairomones: transspecific chemical messengers. *BioScience* **20**, 21-22.
Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics* **81**, 1084-1097.

Bryant EH, McCommas SA, Combs LM (1986) The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. *Genetics* **114**, 1191-1211.
Burge DRL, Edlund MB, Frisch D (2018) Paleolimnology and resurrection ecology: the future of reconstructing the past. *Evol Appl* **11**, 42-59.
Burgess SC, Baskett ML, Grosberg RK, Morgan SG, Strathmann RR (2016) When is dispersal for dispersal? Unifying marine and terrestrial perspectives. *Biol Rev Camb Philos Soc* **91**, 867-882.
Busch R, Qiu W, Lasky-Su J*, et al.* (2016) Differential DNA methylation marks and gene comethylation of COPD in African-Americans with COPD exacerbations. *Respir Res* **17**, 143.
Buskirk JV, Steiner UK (2009) The fitness costs of developmental canalization and plasticity. *J Evol Biol* **22**, 852-860.
Carpenter SR, Cole JJ, Hodgson JR*, et al.* (2001) Trophic cascades, nutrients, and lake productivity: whole-lake experiments. *Ecological Monographs* **71**, 163-186.
Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics* **15**, 97-132.
Castro BB, Consciência S, Gonçalves F (2007) Life history responses of *Daphnia longispina* to mosquitofish (*Gambusia holbrooki)* and pumpkinseed (*Lepomis gibbosus*) kairomones. *Hydrobiologia* **594**, 165-174.
Chalmers KJ, Waugh R, Sprent JI, Simons AJ, Powell W (1992) Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. *Heredity* **69**.
Chapman RW, Mancia A, Beal M*, et al.* (2011) The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular Ecology* **20**, 1431-1449.
Chapman RW, Mancia A, Beal M*, et al.* (2009) A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica,* in the South Atlantic bight. *Molecular Ecology* **18**, 2415-2425.
Charlesworth D, Charlesworth B (1975) Theoretical genetics of batesian mimicry II. Evolution of supergenes. *Journal of Theoretical Biology* **55**, 305-324.
Charmantier A, McCleery RH, Cole LR*, et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**.
Chen HB, Paul C (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**.
Chevin LM, Lande R, Mace GM (2010) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biol* **8**, e1000357.
Cheviron ZA, Whitehead A, Brumfield RT (2008) Transcriptomic variation and plasticity in rufous-collared sparrows (*Zonotrichia capensis)* along an altitudinal gradient. *Molecular Ecology* **17**, 4556-4569.
Cohen GM, Shurin JB (2003) Scale-dependence and mechanisms of dispersal in freshwater zooplankton. *Oikos* **103**, 603-617.
Colbourne JK, Pfrender ME, Gilbert D*, et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555-561.
Colbourne JK, Singan VR, Gilbert DG (2005) wFleaBase: the *Daphnia* genome database. *BMC Bioinformatics* **6**, 45.
Cook DR (1977) Detection of influential observation in linear regression. *Technometrics* **19**, 15-18.
Cousyn C, De Meester L, Colbourne JK*, et al.* (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 of the United States of America **98**, 6256-6260.
Dahl J, Peckarsky BL (2003) Developmental responses to predation risk in morphologically defended mayflies. *Oecologia* **137**, 188-194.
Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
Darwin C (1859) *On the orgin of species*.
Davison J (1969) Activation of the ephippial egg of *Daphnia pulex*. *The Journal of General Physiology*, 562-575.
De Boer ME, Ellers J, van Gestel CA*, et al.* (2013) Transcriptional responses indicate attenuated oxidative stress in the springtail *Folsomia candida* exposed to mixtures of cadmium and phenanthrene. *Ecotoxicology* **22**, 619-631.
de Jong G (2005) Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes. *New Phytol* **166**, 101-117.
De Meester L (1991) An analysis of the phototactic behaviour of *Daphnia magna* clones and their sexual descendants. *Hydrobiologia* **225**, 217-227.
De Meester L (1993) Genotype, fish-mediated chemical, and phototactic behavior in *Daphnia magna*. *Ecology* **74**, 1467-1474.
De Meester L (1996a) Evolutionary potential and local genetic differentiation in a phenotypically plastic trait of a cyclical parthenogen, *Daphnia magna*. *Evolution* **50**, 1293-1298.
De Meester L (1996b) Local genetic differentiation and adaptation in freshwater zooplankton populations: patterns and processes. *EcoScience* **3**, 385-399.
De Meester L, Gomez A, Okamura B, Schwenk K (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica* **23**, 121-135.
De Meester L, Vanoverbeke J, Kilsdonk LJ, Urban MC (2016) Evolving perspectives on monopolization and priority effects. *Trends Ecol Evol* **31**, 136-146.
de Villemereuil P, Mouterde M, Gaggiotti OE, Till-Bottraud I, Jacquemyn H (2018) Patterns of phenotypic plasticity and local adaptation in the wide elevation range of the alpine plant *Arabis alpina*. *Journal of Ecology*.
Declerck S, Cousyn C, De Meester L (2001) Evidence for local adaptation in neighbouring *Daphnia* populations: a laboratory transplant experiment. *Freshwater Biology* **46**, 187-198.
DePristo MA, Banks E, Poplin R*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491-498.
Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Annual Review of Ecology, Evolution, and Systematics* **44**, 5-29.
Desmarais KH (1997) Keeping *Daphnia* out of the surface film with cetyl alcohol. *Journal of Plankton Research* **19**, 149-154.
DeWitt TJ, Scheiner SM (2004) Phenotypic plasticity: functional and conceptual approaches.
DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* **13**.
Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature* **400**.
Dodson SI (1989) The ecological role of chemical stimuli for the zooplankton: predator-induced morphology in *Daphnia*. *Oecologia* **78**, 361-367.

Dodson SI, Hanazato T (1995) Commentary on Effects of Anthropogenic and Natural Organic Chemicals on Development, Swimming Behavior, and Reproduction of *Daphnia*, a Key Member of Aquatic Ecosystems. *Environmental Health Perspectives* **103**, 7-11.
Doebeli M, Blok HJ, Leimar O, Dieckmann U (2007) Multimodal pattern formation in phenotype distributions of sexual populations. *Proc Biol Sci* **274**, 347-357.
Duijvesteijn N, Knol EF, Merks JW*, et al.* (2010) A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genetics* **11**.
Duplouy A, Wong SC, Corander J, Lehtonen R, Hanski I (2017) Genetic effects on life-history traits in the Glanville fritillary butterfly. *PeerJ* **5**, e3371.
Eads BD, Andrews J, Colbourne JK (2008) Ecological genomics in *Daphnia:* stress responses and environmental sex determination. *Heredity (Edinb)* **100**, 184-190.
Ebert D (2005) *Ecology, epidemiology and evolution of parasitism in Daphnia*.
Ebert D, Haag C, Kirkpatrick M*, et al.* (2002) A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science* **295**, 485-488.
Ebert D, Zschokke-Rohringer CD, Carius HJ (1998) Within–and between–population variation for resistance of *Daphnia magna* to the bacterial endoparasite *Pasteuria ramosa*. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 2127-2134.
Edgar R, Domrachev M, Lash AE (2001) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**, 207-210.
Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**, 207-210.
Ehrlich P, Raven P (1969) Differentiation of populations. *Science* **165**, 1228-1232.
Eising E, Huisman SM, Mahfouz A*, et al.* (2016) Gene co-expression analysis identifies brain regions and cell types involved in migraine pathophysiology: a GWAS-based study using the Allen Human Brain Atlas. *Hum Genet* **135**, 425-439.
Estes S, Arnold Stevan J (2007) Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist* **169**, 227-244.
Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Current Biology* **18**, 363-367.
Ferreira A, Marguti I, Bechmann I*, et al.* (2011) Sickle hemoglobin confers tolerance to *Plasmodium* infection. *Cell* **145**, 398-409.
Feulner PG, Chain FJ, Panchal M*, et al.* (2013) Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology* **22**, 635-649.
Filteau M, Pavey SA, St-Cyr J, Bernatchez L (2013) Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol Biol Evol* **30**, 1384-1396.
Finn RD, Bateman A, Clements J*, et al.* (2014) Pfam: the protein families database. *Nucleic Acids Research* **42**, D222-230.
Fisher RA (1930) *The genetical theory of natural selection*.
Fitzpatrick BM (2012) Underappreciated consequences of phenotypic plasticity for ecological speciation. *International Journal of Ecology* **2012**, 1-12.
Fleischer RC, Conant S, Morin MP (1991) Genetic variation in native and translocated populations of the Laysan finch (*Telespiza cantans*). *Heredity* **66**.
Fraser HB (2011) Genome-wide approaches to the study of adaptive gene expression evolution. *Bioessays* **33**, 469-477.

Frisch D, Morton PK, Chowdhury PR, *et al.* (2014) A millennial-scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecol Lett* **17**, 360-368.
Fryer G (1985) The ecology and distribution of the genus *Daphnia* (Crustacea: Cladocera) in restricted areas: the pattern in Yorkshire. *Journal of Natural History* **19**, 97-128.
Fuhrman AE, Larsen DA, Steel EA, Young G, Beckman BR (2018) Chinook salmon emergence phenotypes: describing the relationships between temperature, emergence timing and condition factor in a reaction norm framework. *Ecology of Freshwater Fish* **27**, 350-362.
Fuller SJ, Mather PB, Wilson JC (1996) Limited genetic differentiation among wild *Oryctolagus cuniculus L.* (rabbit) populations in arid eastern Australia. *Heredity* **77**.
Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486-1487.
Fusco G, Minelli A (2010) Phenotypic plasticity in development and evolution: facts and concepts. *Philos Trans R Soc Lond B Biol Sci* **365**, 547-556.
Galesloot TE, van Steen K, Kiemeney LA, Janss LL, Vermeulen SH (2014) A comparison of multivariate genome-wide association methods. *PLoS One* **9**, e95923.
Garbutt JS, Little TJ (2014) Maternal food quantity affects offspring feeding rate in *Daphnia magna*. *Biol Lett* **10**.
Garbutt JS, O'Donoghue AJ, McTaggart SJ, Wilson PJ, Little TJ (2014) The development of pathogen resistance in *Daphnia magna*: implications for disease spread in age-structured populations. *Journal of Experimental Biology* **217**, 3929-3934.
Gavrilets S, Hastings A (1996) Founder effect speciation: a theoretical reassessment. *The American Naturalist* **147**, 466-491.
Geber Monica A, Griffen Lauren R (2003) Inheritance and natural selection on functional traits. *International Journal of Plant Sciences* **164**, S21-S42.
Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology* **21**, 394-407.
Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. *Nat Rev Genet* **5**, 681-690.
Gießler S (2001) Morphological differentiation within the *Daphnia longispina* group. *Hydrobiologia* **442**, 55-66.
Gilbert-Horvath EA, Larson RJ, Garza JC (2006) Temporal recruitment patterns and gene flow in kelp rockfish (*Sebastes atrovirens*). *Mol Ecol* **15**, 3801-3815.
Gilman SE, Urban MC, Tewksbury J, Gilchrist GW, Holt RD (2010) A framework for community interactions under climate change. *Trends Ecol Evol* **25**, 325-331.
Giorello FM, Feijoo M, D'Elía G, *et al.* (2018) An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). *Molecular Ecology*.
Gómez A, Adcock GJ, Lunt DH, Carvalho GR (2002) The interplay between colonization history and gene flow in passively dispersing zooplankton: microsatellite analysis of rotifer resting egg banks. *Journal of Evolutionary Biology* **15**, 158-171.
Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. *Evolution* **42**, 441-454.
Gorr TA, Cahn JD, Yamagata H, Bunn HF (2004) Hypoxia-induced synthesis of hemoglobin in the crustacean *Daphnia magna* is hypoxia-inducible factor-dependent. *J Biol Chem* **279**, 36038-36047.
Grant BR, Grant PR (1989) Natural selection in a population of Darwin's finches. *The American Naturalist* **133**, 377-393.

Grant PR, Grant BR (2006) Evolution of character displacement in Darwin's finches. *Science* **313**, 224-226.
Greene DF (2005) The role of abscission in long-distance seed dispersal by the wind. *Ecology* **86**, 3105-3110.
Grether Gregory F (2005) Environmental change, phenotypic plasticity and genetic compensation. *The American Naturalist* **166**.
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**, 810-823.
Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812.
Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2006) Founder events as determinants of within-island and among-island genetic structure of *Daphnia* metapopulations. *Heredity (Edinb)* **96**, 150-158.
Haileselasie TH, Mergeay J, Vanoverbeke J, Orsini L, De Meester L (2018) Founder effects determine the genetic structure of the water flea *Daphnia* in Ethiopian reservoirs. *Limnology and Oceanography* **63**, 915-926.
Hairston Jr NG, Lampert W, CÃ¡ceres CE*, et al.* (1999) Rapid evolution revealed by dormant eggs. *Nature* **401**.
Hamrova E, Mergeay J, Petrusek A (2011) Strong differences in the clonal variation of two *Daphnia* species from mountain lakes affected by overwintering strategy. *BMC Evolutionary Biology* **11**.
Hansen MM, Olivieri I, Waller DM, Nielsen EE, Ge MWG (2012) Monitoring adaptive genetic responses to environmental change. *Mol Ecol* **21**, 1311-1329.
Harvell CD (1990) The ecology and evolution of inducible defenses. *The Quarterly Review of Biology* **65**, 323-340.
Haupt F, Stockenreiter M, Baumgartner M, Boersma M, Stibor H (2009) *Daphnia* diel vertical migration: implications beyond zooplankton. *Journal of Plankton Research* **31**, 515-524.
Havel JE (2004) Mechanisms, effects, and scales of dispersal in freshwater zooplankton.
Havel JE, Colbourne JK, Hebert PDN (2000) Reconstructing the history of intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers. *Limnology and Oceanography* **45**, 1414-1419.
Hayashi Y, Heckmann LH, Callaghan A, Sibly RM (2008) Reproduction recovery of the crustacean *Daphnia magna* after chronic exposure to ibuprofen. *Ecotoxicology* **17**, 246-251.
Hayes BJ, Bowman PJ, Chamberlain AJ*, et al.* (2009) A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS One* **4**, e6676.
Heather GP, Colbourne JK, Wendel BK (2003) Reconstruction of Centuries-Old *Daphnia* Communities in a Lake Recovering from Acidification and Metal Contamination. *Ambio* **32**, 214-218.
Hebert PDN (1978) The population biology of *Daphnia* (Crustacea, Daphnidae). *Biological Reviews* **53**, 387-426.
Hebert PDN, Ward RD (1972) Inheritance during parthenogenesis in *Daphnia magna*. *Genetics* **71**, 639-642.
Heckmann LH, Callaghan A, Hooper HL*, et al.* (2007) Chronic toxicity of ibuprofen to *Daphnia magna*: effects on life history traits and population dynamics. *Toxicology Letters* **172**, 137-145.

Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology, Evolution, and Systematics* **37**, 67-93.
Hendry AP, Kinnison MT (1999) Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**, 1637-1653.
Henning-Lucass N, Cordellier M, Streit B, Schwenk K (2016) Phenotypic plasticity in life-history traits of *Daphnia galeata* in response to temperature - a comparison across clonal lineages separated in time. *Ecology and Evolution* **6**, 881-891.
Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist* **173**, 579-588.
Hereford J, Hansen TF, Houle D (2007) Comparing strengths of directional selection: how strong is strong? *Evolution* **58**, 2´133-2143.
Hermisson J, Wagner GP (2004) The population genetic theory of hidden variation and genetic robustness. *Genetics* **168**, 2271-2284.
Herrmann M (2017) *Evolutionary Genomics in Daphnia*, Universität Koblenz-Landau.
Herrmann M, Henning-Lucass N, Cordellier M, Schwenk K (2017a) A genotype-phenotype association approach to reveal thermal adaptation in *Daphnia galeata*. *J Exp Zool A Ecol Integr Physiol* **327**, 53-65.
Herrmann M, Ravindran SP, Schwenk K, Cordellier M (2017b) Data from: Population transcriptomics in *Daphnia*: the role of thermal selection. Dryad Data Repository.
Herrmann M, Ravindran SP, Schwenk K, Cordellier M (2017c) Population transcriptomics in *Daphnia*: the role of thermal selection. *Molecular Ecology*.
Hill WG, Barton NH, Turelli M (2006) Prediction of effects of genetic drift on variance components under a general model of epistasis. *Theoretical Population Biology* **70**, 56-62.
Hoban S, Kelley JL, Lotterhos KE*, et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions and future directions. *Am Nat* **188**, 379-397.
Hodgins KA, Yeaman S, Nurkowski KA, Rieseberg LH, Aitken SN (2016) Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Molecular Biology and Evolution* **33**, 1502-1516.
Hoelzel AR (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society* **68**, 23-39.
Hoelzel AR, Fleischer RC, Campagna C, Le Boeuf BJ, Alvord G (2002) Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology* **15**, 567-575.
Hoffmann AA, Hallas RJ, Dean JA, Schiffer M (2003) Low potential for climatic stress adaptation in a rainforest *Drosophila* Species. *Science* **301**.
Houck MA, OConnor BM (1991) Ecological and evolutionary significance of phoresy in the Astigmata. *Annual Review of Entomology* **36**, 611-636.
Husby A, Kawakami T, Ronnegard L*, et al.* (2015) Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proc Biol Sci* **282**, 20150156.
Huylmans AK, Lopez 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**, 3120-3139.
Jansen RC, Nap J-P (2001) Genetical genomics: the added value from segregation. *Trends in Genetics* **17**, 388-391.

Janzen FJ, Phillips PC (2006) Exploring the evolution of environmental sex determination, especially in reptiles. *J Evol Biol* **19**, 1775-1784.
Jin W, Riley RM, Wolfinger RD*, et al.* (2001) The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. *Nat Genet* **29**, 389-395.
Joachim M, Joost V, Dirk V, De ML (2007) Extinction, recolonization, and dispersal through time in a planktonic crustacean.
Johnson PTJ, Olden JD, Vander Zanden MJ (2008) Dam invaders: impoundments facilitate biological invasions into freshwaters. *Frontiers in Ecology and the Environment* **6**, 357-363.
Johnson T, Barton N (2005) Theoretical models of selection and mutation on quantitative traits. *Philos Trans R Soc Lond B Biol Sci* **360**, 1411-1425.
Johnson TC, Scholz CA, Talbot MR*, et al.* (1996) Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**, 1091-1093.
Johnston SE, Gratten J, Berenos C*, et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.
Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* **374**, 20150202.
Jueterbock A, Franssen SU, Bergmann N*, et al.* (2016) Phylogeographic differentiation versus transcriptomic adaptation to warm temperatures in *Zostera marina*, a globally important seagrass. *Molecular Ecology* **25**, 5396-5411.
Kao PY, Leung KH, Chan LW, Yip SP, Yap MK (2017) Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. *Biochim Biophys Acta* **1861**, 335-353.
Karban RaB, I.T., (1989) Induced plant responses to herbivory. *Annual Review of Ecology and Systematics* **20**, 331-348.
Karhunen M, Merila J, Leinonen T, Cano JM, Ovaskainen O (2013) DRIFTSEL: an R package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources* **13**, 746-754.
Karhunen M, Ovaskainen O (2012) Estimating population-level coancestry coefficients by an admixture F model. *Genetics* **192**, 609-617.
Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.
Kawecki TJ (1994) Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *The American Naturalist* **144**, 833-838.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
Keller B, Wolinska J, Manca M, Spaak P (2008) Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philos Trans R Soc Lond B Biol Sci* **363**, 2943-2952.
Keller LF, Arcese P, Smith JNM, Hochachka WM, Stearns SC (1994) Selection against inbred song sparrows during a natural population bottleneck. *Nature* **372**.
Keller LFW, Donald M. (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**, 230-241.
Kellermann V, van Heerwaarden B, Sgrò CM, Hoffmann AA (2009) Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* **325**, 1244-1246.
Kerfoot WC, Robbins JA, Weider LJ (1999) A new approach to historical reconstruction: Combining descriptive and experimental paleolimnology. *Limnology and Oceanography* **44**, 1232-1247.

Kesari R, Lasky JR, Villamor JG, *et al.* (2012) Intron-mediated alternative splicing of *Arabidopsis P5CS1* and its association with natural variation in proline and climate adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9197-9202.
Kim S-C, Rieseberg LH (1999) Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics* **153**, 965-977.
Kingsolver JG, Diamond SE (2011) Phenotypic selection in natural populations: what limits directional selection? *Am Nat* **177**, 346-357.
Kingsolver JG, Hoekstra HE, Hoekstra JM, *et al.* (2001) The strength of phenotypic selection in natural populations. *The American Naturalist* **157**, 245-261.
Kishida O, Trussell GC, Mougi A, Nishimura K (2009) Evolutionary ecology of inducible morphological plasticity in predator–prey interaction: toward the practical links with population ecology. *Population Ecology* **52**, 37-46.
Kitano J, Bolnick DI, Beauchamp DA, *et al.* (2008) Reverse evolution of armor plates in the threespine stickleback. *Current Biology* **18**, 769-774.
Klein D, Ono H, O'hUigin C, *et al.* (1993) Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature* **364**.
Koehn RK, Diehl WJ, Scott TM (1988) The differential contribution by individual enzymes of glycolysis and protein catabolism to the relationship between heterozygosity and growth rate in the coot clam, *Mulinia lateralis*. *Genetics* **118**, 121-130.
Korpelainen H (1986) Competition between clones: an experimental study in a natural population of *Daphnia magna*. *Hereditas* **105**, 29-35.
Kudapa H, Azam S, Sharpe AG, *et al.* (2014) Comprehensive transcriptome assembly of chickpea (*Cicer arietinum L.*) using Sanger and next generation sequencing platforms: development and applications. *PLoS One* **9**, e86039.
Lafferty KD (2009) The ecology of climate change and infectious diseases. *Ecology*.
Laforsch C, Beccara L, Tollrian R (2006) Inducible defenses: the relevance of chemical alarm cues in *Daphnia*. *Limnology and Oceanography* **51**, 1466-1472.
Laforsch C, Tollrian R (2004a) Extreme helmet formation in *Daphnia cucullata* induced by small-scale turbulence. *Journal of Plankton Research* **26**, 81-87.
Laforsch C, Tollrian R (2004b) Inducible defenses in multipredator environments: cyclomorphosis in *Daphnia cucullata*. *Ecology* **85**, 2302-2311.
Laforsch C, Tollrian R (2009) Cyclomorphosis and phenotypic changes.
Lahti DC, Johnson NA, Ajie BC, *et al.* (2009) Relaxed selection in the wild. *Trends Ecol Evol* **24**, 487-496.
Lampert W (1993) Phenotypic plasticity of the size at first reproduction in *Daphnia:* the importance of maternal size. *Ecology* **74**, 1455-1466.
Lampert W, Lampert KP, Larsson P (2012) Induction of male production in clones of *Daphnia pulex* by the juvenoid hormone methyl farnesoate under short photoperiod. *Comp Biochem Physiol C Toxicol Pharmacol* **156**, 130-133.
Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* **37**, 1210-1226.
Langerhans RB, Dewitt TJ (2002) Plasticity constrained: over-generalized induction cues cause maladaptive phenotypes.
Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? *PLoS Comput Biol* **7**, e1001057.

Larsen PF, Nielsen EE, Williams TD, Loeschcke V (2008) Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*). *Heredity (Edinb)* **101**, 247-259.
Lass S, Spaak P (2003) Chemically induced anti-predator defences in plankton: a review. *Hydrobiologia* **491**, 221-239.
Le Rouzic A, Carlborg O (2008) Evolutionary potential of hidden genetic variation. *Trends Ecol Evol* **23**, 33-37.
Leary RF, Allendorf FW, Knudsen KL, Thorgaard GH (1985) Heterozygosity and developmental stability in gynogenetic diploid and triploid rainbow trout. *Heredity* **54**.
Lemon J (2006) Plotrix: a package in the red light district of R. **6**, 8-12.
Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**, 2178-2189.
Lind MI, Ingvarsson PK, Johansson H, Hall D, Johansson F (2011) Gene flow and selection on phenotypic plasticity in an island system of *Rana temporaria*. *Evolution* **65**, 684-697.
Llaurens V, Whibley A, Joron M (2017) Genetic architecture and balancing selection: the life and death of differentiated variants. *Mol Ecol* **26**, 2430-2448.
Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics* **9**, 583-593.
Louette G, De Meester L (2004) Rapid colonization of a newly created habitat by cladocerans and the initial build-up of a *Daphnia*-dominated community. *Hydrobiologia* **513**, 245-249.
Louette G, De Meester L (2005) High dispersal capacity of cladoceran zooplankton in newly founded communities. *Ecology* **86**, 353-359.
Louette G, Vanoverbeke J, Ortells R, De Meester L (2007) The founding mothers: the genetic structure of newly established *Daphnia* populations. *Oikos* **116**, 728-741.
Love MI, Huber W, Anders S (2014) Differential analysis of count data – the DESeq2 package. *bioRxiv (2014)*.
Luikart G, Allendorf F, Cornuet J, Sherwin W (1998a) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.
Luikart GA, FW, Cornuet J-M, Sherwin W (1998b) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.
Lynch MG, Wilfried (1983) Phenotypic evolution and parthenogenesis. *The American Naturalist* **122**, 745-764.
Macháček J (1991) Indirect effect of planktivorous fish on the growth and reproduction of *Daphnia galeata*. *Hydrobiologia* **225**, 193-197.
Macháček J (1995) Inducibility of life history changes by fish kairomone in various developmental stages of *Daphnia*. *Journal of Plankton Research* **17**, 1513-1520.
MacManes MD, Eisen MB (2014) Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus*. *PeerJ* **2**, e642.
Macnair MR (2002) Within and between population genetic variation for zinc accumulation in *Arabidopsis halleri*. *New Phytologist* **155**, 59-66.
Matthysen E (2005) Density-dependent dispersal in birds and mammals. *Ecography* **28**, 403-416.

Matute DR (2013) The role of founder effects on the evolution of reproductive isolation. *J Evol Biol* **26**, 2299-2311.
McGuigan K, Sgro CM (2009) Evolutionary consequences of cryptic genetic variation. *Trends Ecol Evol* **24**, 305-311.
McKenna A, Hanna M, Banks E*, et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303.
McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution* **17**, 480-488.
McKown AD, Klapste J, Guy RD*, et al.* (2014) Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytol* **203**, 535-553.
McMillian M, Nie AY, Parker JB*, et al.* (2004) A gene expression signature for oxidant stress/reactive metabolites in rat liver. *Biochem Pharmacol* **68**, 2249-2261.
Mellors KW (1975) Selective predation of ephippal *Daphnia* and the resistance of ephippal eggs to digestion. *Ecology* **56**, 974-980.
Menken SBJ (1987) Is the extremely low heterozygosity level in *Yponomeuta rorellus* caused by bottlenecks? *Evolution* **41**, 630-637.
Mergeay J, Aguilera X, Declerck S*, et al.* (2008) The genetic legacy of polyploid Bolivian *Daphnia:* the tropical Andes as a source for the North and South American *D. pulicaria* complex. *Molecular Ecology* **17**, 1789-1800.
Merila J (2012) Evolution in response to climate change: in pursuit of the missing evidence. *Bioessays* **34**, 811-818.
Merila J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evol Appl* **7**, 1-14.
Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**.
Mills S, Lunt DH, Gomez A (2007) Global isolation by distance despite strong regional phylogeography in a small metazoan. *BMC Evolutionary Biology* **7**, 225.
Miner BE, De Meester L, Pfrender ME, Lampert W, Hairston NG, Jr. (2012) Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings. Biological sciences* **279**, 1873-1882.
Miner BG, Sultan SE, Morgan SG, Padilla DK, Relyea RA (2005) Ecological consequences of phenotypic plasticity. *Trends Ecol Evol* **20**, 685-692.
Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* **41**, e121.
Mitton JB (1978) Relationship between heterozygosity for enzyme loci and variation of morphological characters in natural populations. *Nature* **273**.
Moczek AP (2008) On the origins of novelty in development and evolution. *Bioessays* **30**, 432-447.
Morgan KK, Hicks J, Spitze K*, et al.* (2007) Patterns of genetic architecture for life-history traits and molecular markers in a subdivided species. *Evolution* **55**, 1753-1761.
Mousseau TA, Roff DA (1987) Natural selection and the heritability of fitness components. *Heredity* **59**.
Munoz J, Chaturvedi A, De Meester L, Weider LJ (2016) Characterization of genome-wide SNPs for the water flea *Daphnia pulicaria* generated by genotyping-by-sequencing (GBS). *Scientific Reports* **6**, 28569.

Munoz J, Gomez A, Green AJ*, et al.* (2008) Phylogeography and local endemism of the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda: Anostraca). *Molecular Ecology* **17**, 3160-3177.
Myers SS, Gaffikin L, Golden CD*, et al.* (2013) Human health impacts of ecosystem alteration. *Proc Natl Acad Sci U S A* **110**, 18753-18760.
Naihong X, Audenaert E, Vanoverbeke J*, et al.* (2000) Low among-population genetic differentiation in Chinese bisexual *Artemia* populations. *Heredity* **84**.
Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics* **76**, 379-390.
Okun N, Mendonca R, Mehner T (2005) Diel shifts in community composition and feeding of juvenile fishes in the pelagic area of a large shallow lake. *Limnologica - Ecology and Management of Inland Waters* **35**, 70-77.
Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics* **32**, 261-266.
Olmstead AW, LeBlanc GA (2003) Insecticidal Juvenile Hormone Analogs Stimulate the Production of Male Offspring in the Crustacean *Daphnia magna*. *Environmental Health Perspectives* **111**, 919-924.
Orsini L, Brown JB, Shams Solari O*, et al.* (2017) Early transcriptional response pathways in *Daphnia magna* are coordinated in networks of crustacean-specific genes. *Mol Ecol*.
Orsini L, Gilbert D, Podicheti R*, et al.* (2016) *Daphnia magna* transcriptome by RNA-Seq across 12 environmental stressors. *Scientific Data* **3**, 160030.
Orsini L, Schwenk K, De Meester L*, et al.* (2013a) The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. *Trends Ecol Evol* **28**, 274-282.
Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013b) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**, 5983-5999.
Ortells R, Vanoverbeke J, Louette G, De Meester L (2013) Colonization of *Daphnia magna* in a newly created pond: founder effects and secondary immigrants. *Hydrobiologia* **723**, 167-179.
Östman Ö (2011) Interspecific competition affects genetic structure but not genetic diversity of *Daphnia magna*. *Ecosphere* **2**.
Paaby AB, Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate. *Nat Rev Genet* **15**, 247-258.
Pai AA, Pritchard JK, Gilad Y (2015) The genetic and mechanistic basis for variation in gene regulation. *PLoS Genetics* **11**.
Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419-420.
Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290.
Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods Ecol Evol* **6**, 445-464.
Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537-2539.
Petrusek A, Černý M, Mergeay J, Schwenk K (2007) *Daphnia* in the Tatra Mountain lakes: multiple colonisation and hidden species diversity revealed by molecular markers. *Fundamental and Applied Limnology / Archiv für Hydrobiologie* **169**, 279-291.

Petryszak R, Keays M, Tang YA, *et al.* (2016) Expression Atlas update--an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res* **44**, D746-752.
Pfennig DW, Rice AM, Martin RA (2007) Field and experimental evidence for competition's role in phenotypic divergence. *Evolution* **61**, 257-271.
Pfennig DW, Wund MA, Snell-Rood EC, *et al.* (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol* **25**, 459-467.
Pietrzak B, Pijanowska J, Dawidowicz P (2015) The effect of temperature and kairomone on *Daphnia* escape ability: a simple bioassay. *Hydrobiologia* **798**, 15-23.
Pietrzak BS, Miroslaw (2006) The fate of the ephippia-*Daphnia* dispersal in time and space. *Polish Journal of Ecology* **54**, 709-714.
Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends Ecol Evol* **20**, 481-486.
Pigliucci M, Murren CJ, Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* **209**, 2362-2367.
Pijanowska J (1997) Alarm signals in *Daphnia? Oecologia* **112**, 12-16.
Pijanowska J, Kowalczewski A (1997) Cues from injured *Daphnia* and from cyclopoids feeding on *Daphnia* can modify life histories of conspecifics. *Hydrobiologia* **350**, 99-103.
Pohnert G, Steinke M, Tollrian R (2007) Chemical cues, defence metabolites and the shaping of pelagic interspecific interactions. *Trends Ecol Evol* **22**, 198-204.
Poynton HC, Zuzow R, Loguinov AV, Perkins EJ, Vulpe CD (2008) Gene expression profiling in *Daphnia magna,* part II: validation of a copper specific gene expression signature with effluent from two copper mines in California. *Environmental Science & Technology* **42**, 6257-6263.
Pritchard CC, Hsu L, Delrow J, Nelson PS (2001) Project normal: Defining normal variance in mouse gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13266-13271.
Priyam A, Woodcroft BJ, Rai V, *et al.* (2015) Sequenceserver: a modern graphical user interface for custom BLAST databases. *biorxiv*.
Proctor WV (1964) Viability of crustacean eggs recovered from ducks. *Ecology* **45**, 656-658.
Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575.
Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. *Nature* **417**.
R Development Core Team (2008) R: A language and environment for statistical computing.
Radwan J, Babik W (2012) The genomics of adaptation. *Proc Biol Sci* **279**, 5024-5028.
Rafferty NE, Ives AR (2011) Effects of experimental shifts in flowering phenology on plant-pollinator interactions. *Ecol Lett* **14**, 69-74.
Rausher MD, Delph LF (2015) Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution* **69**, 1655-1664.
Rengefors K, Karlsson I, Hansson LA (1998) Algal cyst dormancy: a temporal escape from herbivory. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 1353-1358.
Reusch TB, Veron AS, Preuss C, *et al.* (2008) Comparative analysis of expressed sequence tag (EST) libraries in the seagrass *Zostera marina* subjected to temperature stress. *Marine Biotechnology* **10**, 297-309.

Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112**, 183-198.
Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. **16**.
Roberge C, Guderley H, Bernatchez L (2007) Genomewide identification of genes under directional selection: gene transcription Q_{ST} scan in diverging Atlantic salmon subpopulations. *Genetics* **177**, 1011-1022.
Robertson A (1952) The effect of inbreeding on the variation due to recessive genes. *Genetics* **37**, 189-207.
Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* **456**, 738-744.
Roelofs D, Janssens TK, Timmermans MJ*, et al.* (2009) Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod *Orchesella cincta*. *Molecular Ecology* **18**, 3227-3239.
Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* **13**, 505-516.
Ronnegard L, McFarlane SE, Husby A*, et al.* (2016) Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation. *Methods Ecol Evol* **7**, 792-799.
Roslin T (2001) Spatial population structure in a patchily distributed beetle. *Molecular Ecology* **10**, 823-837.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
Routtu J, Ebert D (2015) Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasites. *Heredity (Edinb)* **114**, 241-248.
Routtu J, Hall MD, Albere B*, et al.* (2014) An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. *BMC Genomics* **15**.
Roy Chowdhury P, Frisch D, Becker D*, et al.* (2015) Differential transcriptomic responses of ancient and modern *Daphnia* genotypes to phosphorus supply. *Molecular Ecology* **24**, 123-135.
Roy Chowdhury P, Lopez JA, Weider LJ, Colbourne JK, Jeyasingh PD (2014) Functional genomics of intraspecific variation in carbon and phosphorus kinetics in *Daphnia*. *Journal of Experimental Zoology. Part A, Ecological Genetics and Physiology* **321**, 387-398.
Rueffler C, Van Dooren TJ, Leimar O, Abrams PA (2006) Disruptive selection and then what? *Trends Ecol Evol* **21**, 238-245.
Rundle HD, Mooers AØ, Whitlock MC (2017) Single founder-flush events and the evolution of reproductive isolation. *Evolution* **52**, 1850-1855.
Saastamoinen M, Bocedi G, Cote J*, et al.* (2018) Genetics of dispersal. *Biol Rev Camb Philos Soc* **93**, 574-599.
Saccheri I, Kuussaari M, Kankare M*, et al.* (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**.
Sakwińska O (2008) Plasticity of *Daphnia magna* life history traits in response to temperature and information about a predator. *Freshwater Biology* **39**, 681-687.
Salamin N, Wuest RO, Lavergne S, Thuiller W, Pearman PB (2010) Assessing rapid evolution in a changing environment. *Trends Ecol Evol* **25**, 692-698.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**, 807-820.
Schadt EE, Monks SA, Drake TA*, et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302.
Schluter D (1994) Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798-801.
Schluter D (2003) Frequency dependent natural selection during character displacement in sticklebacks. *Evolution* **57**, 1142-1150.
Schwenk K (1993) Interspecific hybridization in *Daphnia*: distinction and origin of hybrid matrilines. *Molecular Biology and Evolution* **10**, 1289-1302.
Schwenk K, Posada D, Hebert PD (2000) Molecular systematics of European *Hyalodaphnia*: the role of contemporary hybridization in ancient species. *Proc Biol Sci* **267**, 1833-1842.
Schwerin S, Zeis B, Lamkemeyer T*, et al.* (2009) Acclimatory responses of the *Daphnia pulex* proteome to environmental changes. II. Chronic exposure to different temperatures (10 and 20 degrees C) mainly affects protein metabolism. *BMC Physiol* **9**, 8.
Sedlazeck FJ, Rescheneder P, von Haeseler A (2013) NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791.
Sharma A, Sharma VK, Horn-Saban S*, et al.* (2005) Assessing natural variations in gene expression in humans by comparing with monozygotic twins using microarrays. *Physiol Genomics* **21**, 117-123.
Shaw JR, Pfrender ME, Eads BD*, et al.* (2008) *Daphnia* as an emerging model for toxicological genomics. In: *Comparative Toxicogenomics*, pp. 165-328.
Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecol Lett* **12**, 1261-1276.
Silberbush A, Markman S, Lewinsohn E*, et al.* (2010) Predator-released hydrocarbons repel oviposition by a mosquito. *Ecol Lett* **13**, 1129-1138.
Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology, Evolution and Systematics* **16**, 393-430.
Slatkin M (2004) A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *American Journal of Human Genetics* **75**, 282-293.
Smarda P, Horova L, Bures P, Hralova I, Markova M (2010) Stabilizing selection on genome size in a population of *Festuca pallens* under conditions of intensive intraspecific competition. *New Phytol* **187**, 1195-1204.
Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. *PLoS Biol* **6**, e83.
Smith TB (1993) Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes*. *Nature* **363**.
Snæbjörn P (2001) Microsatellite variation in *Daphnia pulex* from both sides of the Baltic Sea. *Molecular Ecology* **9**, 1075-1088.
Snell-Rood EC, Cash A, Han MV*, et al.* (2011) Developmental decoupling of alternative phenotypes: insights from the transcriptomes of horn-polyphenic beetles. *Evolution* **65**, 231-245.
Snell-Rood EC, Van Dyken JD, Cruickshank T, Wade MJ, Moczek AP (2010) Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays* **32**, 71-81.
Soetaert A, Vandenbrouck T, van der Ven K*, et al.* (2007) Molecular responses during cadmium-induced stress in *Daphnia magna*: integration of differential gene expression with higher-level effects. *Aquatic Toxicology* **83**, 212-222.

Sommer U, Sommer F, Santer B*, et al.* (2003) *Daphnia* versus copepod impact on summer phytoplankton: functional compensation at both trophic levels. *Oecologia* **135**, 639-647.
Soons MB, Heil GW, Nathan R, Katul GG (2004) Determinants of long-distance seed dispersal by wind in grasslands. *Ecology* **85**, 3056-3068.
Soulé ME (1979) Heterozygosity and developmental stability: another look. *Evolution* **33**, 396-401.
Spitze K (1993) Population structure in *Daphnia obtusa* quantitative genetic and allozymic variation. *Genetics*, 367-374.
St Pierre SE, Ponting L, Stefancsik R, McQuilton P, FlyBase C (2014) FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Research* **42**, D780-788.
Stabell OB, Ogbebo F, Primicerio R (2003) Inducible defences in *Daphnia* depend on latent alarm signals from conspecific prey activated in predators. *Chemical Senses* **28**, 141-153.
Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection in human history. *Mol Biol Evol* **22**, 63-73.
Stearns S (1989) The evolutionary significance of phenotypic plasticity. *BioScience* **39**, 436-445.
Steinberg CEW (2012) *Stress Ecology* Springer, Dordrecht.
Stevens VM, Trochet A, Van Dyck H, Clobert J, Baguette M (2012) How is dispersal integrated in life histories: a quantitative analysis using butterflies. *Ecol Lett* **15**, 74-86.
Stevens VM, Whitmee S, Le Galliard JF*, et al.* (2014) A comparative analysis of dispersal syndromes in terrestrial and semi-terrestrial animals. *Ecol Lett* **17**, 1039-1052.
Stibor H (1992) Predator induced life-history shifts in a freshwater cladoceran. *Oecologia* **92**, 162-165.
Stibor H, Lüning J (1994) Predator-induced phenotypic variation in the pattern of growth and reproduction in *Daphnia hyalina* (Crustacea: Cladocera). *Functional Ecology* **8**, 97-101.
Stillman JH, Colbourne JK, Lee CE*, et al.* (2008) Recent advances in crustacean genomics. *Integr Comp Biol* **48**, 852-868.
Straile DG, Walter (1998) Crustacean zooplankton in Lake Constance from 1920 to 1995: Response to eutrophication and re-oligotrophication. *Advances in Limnology* **53**, 255-274.
Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367-383.
Subramanian A, Tamayo P, Mootha VK*, et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545-15550.
Sudheesh S, Sawbridge TI, Cogan NO*, et al.* (2015) *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics* **16**, 611.
Sultan SE (2000) Phenotypic plasticity for plant development, function and life history. *Trends Plant Sci* **5**, 537-542.
Taillon Miller P, Gu Z, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequence a treasure trove of single nucleotide polymorphisms. *Genome Research* **8**, 748-754.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
Tak YG, Farnham PJ (2015) Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57.

Tams V, Luneburg J, Seddar L, Detampel JP, Cordellier M (2018) Intraspecific phenotypic variation in life history traits of Daphnia galeata populations in response to fish kairomones. *PeerJ* **6**, e5746.
Taylor DJ, Hebert PDN, Colbourne JK (1996) Phylogenetics and evolution of the *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and allozyme variation. *Molecular Phylogenetics and Evolution* **5**, 495-510.
Templeton AR (1979) The theory of speciation via the founder principle. *Genetics* **94**, 1011-1038.
Tessier AJ, Leibold MA, Tsao J A fundamental trade-off in resource exploitation by *Daphnia* and consequences to plankton communities. *Ecology* **81**, 826-841.
Thibert-Plante X, Hendry AP (2011) The consequences of phenotypic plasticity for ecological speciation. *J Evol Biol* **24**, 326-342.
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**, 1616-1628.
Thielsch A, Glass N, Streit B*, et al.* (2015) Fitness differences and persistent founder effects determine the clonal composition during population build-up in *Daphnia*. *Oikos* **124**, 620-628.
Thomson FJ, Moles AT, Auld TD*, et al.* (2010) Chasing the unknown: predicting seed dispersal mechanisms from plant traits. *Journal of Ecology* **98**, 1310-1318.
Tollrian R (1995) Predator-induced morphological defenses: costs, life history shifts, and maternal effects in *Daphnia pulex*. *Ecology* **76**, 1691-1705.
Toth GB, Noren F, Selander E, Pavia H (2004) Marine dinoflagellates show induced life-history shifts to escape parasite infection in response to water-borne signals. *Proc Biol Sci* **271**, 733-738.
Travis J (1989) The role of optimizing selection in natural populations. *Annual Review of Ecology and Systematics* **20**, 279-296.
Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology & Evolution* **16**, 330-343.
Valladares F, Matesanz S, Guilhaumon F*, et al.* (2014) The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. *Ecol Lett* **17**, 1351-1364.
van de Meutter F, Stoks R, de Meester L (2008) Size-selective dispersal of *Daphnia* resting eggs by backswimmers (*Notonecta maculata*). *Biol Lett* **4**, 494-496.
Van Dyck H, Baguette M (2005) Dispersal behaviour in fragmented landscapes: routine or special movements? *Basic and Applied Ecology* **6**, 535-545.
Van Dyken JD, Wade MJ (2010) The genetic signature of conditional expression. *Genetics* **184**, 557-570.
van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. *Proc Biol Sci* **276**, 657-665.
Vanoverbeke J, De Meester L (1997a) Among-populational genetic differentiation in the cyclical parthenogen *Daphnia magna* (Crustacea, Anomopoda) and its relation to geographic distance and clonal diversity. *Hydrobiologia* **360**, 135-142.
Vanoverbeke J, De Meester L (1997b) Among-populational genetic differentiation in the cyclical parthenogen *Daphnia magna* (Crustacea, Anomopoda) and its relation to geographic distance and clonal diversity. *Hydrobiologia* **360**, 135-142.
Vatanparast M, Shetty P, Chopra R*, et al.* (2016) Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). *Scientific Reports* **6**, 29070.

Veeramah KR, Hammer MF (2014) The impact of whole-genome sequencing on the reconstruction of human population history. *Nat Rev Genet* **15**, 149-162.
Ventura M, Petrusek A, Miro A*, et al.* (2014) Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential. *Molecular Ecology* **23**, 1014-1027.
Vincek V, O'Huigin C, Satta Y*, et al.* (1997) How large was the founding population of Darwin's finches? *Proceedings of the Royal Society B: Biological Sciences* **264**, 111-118.
Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* **150**.
Waddington CH (1957) *The strategy of the genes*.
Walsh B, Blows MW (2009) Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annual Review of Ecology, Evolution, and Systematics* **40**, 41-59.
Walsh HE, Jones IL, Friesen VL (2005) A test of founder effect speciation using multiple loci in the auklets (*Aethia spp.*). *Genetics* **171**, 1885-1894.
Walther GR (2010) Community and ecosystem responses to recent climate change. *Philos Trans R Soc Lond B Biol Sci* **365**, 2019-2024.
Wang J, Caballero A, Keightley PD, Hill WG (1998) Bottleneck effect on genetic variance: a theoretical investigation of the role of dominance. *Genetics* **150**, 435-447.
Weisberg SF, J (2011) *An R Companion to Applied Regression* Thousand Oaks: Sage.
Weiss L, Laforsch C, Tollrian R (2012) The taste of predation and the defences of prey. In: *Chemical Ecology in Aquatic Systems*, pp. 111-126.
Weiss LC, Leimann J, Tollrian R (2015) Predator-induced defences in *Daphnia longicephala:* location of kairomone receptors and timeline of sensitive phases to trait formation. *J Exp Biol* **218**, 2918-2926.
Werren JH, Richards S, Desjardins CA*, et al.* (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**, 343-348.
Whalley HJ, Knight MR (2012) Calcium signatures are decoded by plants to give specific gene responses. *New Phytologist* **197**, 690-693.
Whitehead A, Crawford DL (2006a) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5425-5430.
Whitehead A, Crawford DL (2006b) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology* **15**, 1197-1211.
Wickham H (2009) ggplot2.
Willi Y, Van Buskirk J, Schmid B, Fischer M (2007) Genetic isolation of fragmented populations is exacerbated by drift and selection. *J Evol Biol* **20**, 534-542.
Willis JH, Orr HA (1993) Increased heritable variation following population bottlenecks: the role of dominance. *Evolution* **47**, 949-957.
Wood SH, Ke X, Nuttall T*, et al.* (2009) Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. *Immunogenetics* **61**, 765-772.
Wood TB, A.M. (1933) Observations on procuring and hatching sexual eggs of *Daphnia longispina.*, 437-454.
Wray GA, Hahn MW, Abouheif E*, et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-1419.
Yampolsky LY, Zeng E, Lopez J*, et al.* (2014) Functional genomics of acclimation and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics* **15**.

Yang LH, Rudolf VH (2010) Phenology, ontogeny and the effects of climate change on the timing of species interactions. *Ecol Lett* **13**, 1-10.
Yasumoto K, Nishigami A, Aoi H*, et al.* (2008) Isolation and absolute configuration determination of aliphatic sulfates as the *Daphnia* kairomones inducing morphological defense of a phytoplankton- part 2. *Chemical and Pharmacuetical Bulletin* **56**, 129-132.
Yasumoto K, Nishigami A, Kasai F, Kusumi T, Ooi T (2006) Isolation and absolute configuration determination of aliphatic sulfates as the *Daphnia* kairomones inducing morphological defense of a phytoplankton. *Chemical and Pharmaceutical Bulletin* **54**, 271-274.
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**.
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**, 4168-4178.
Young TPea (2003) Effects of natural and simulated herbivory on spine lengths of *Acacia drepanolobium* in Kenya. *Oikos* **101**, 171-179.
Zhao K, Tung CW, Eizenga GC*, et al.* (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* **2**, 467.
Zheng X, Levine D, Shen J*, et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326-3328.
Ziv NaS, Bentley M. and Siegal, Mark L. and Gresham, David (2017) Resolving the complex genetic basis of phenotypic variation and variability of cellular growth. *Genetics* **206**.