Evolutionary Genomics in *Daphnia*

by

Maike Herrmann

from Frankfurt am Main

Accepted Dissertation thesis for the partial fulfilment of the requirements for a

Doctor of Natural Sciences

Fachbereich 7: Natur- und Umweltwissenschaften

Universität Koblenz-Landau

Thesis examiners:

Prof. Dr. Klaus Schwenk , Universität Koblenz-Landau

Prof. Dr. Mathilde Cordellier, Universität Hamburg

Prof. Dr. Klaus Fischer, Ernst-Moritz-Arndt-Universität Greifswald

Date of the oral examination: 26. June 2017

"Crawling at your feet," said the Gnat … "you may observe a Bread-and-Butterfly. …"

"And what does it live on?"

"Weak tea with cream in it."

A new difficulty came into Alice's head. "Supposing it couldn't find any?" she suggested.

"Then it would die, of course."

"But that must happen very often," Alice remarked thoughtfully.

"It always happens," said the Gnat.

Lewis Carroll, Through the Looking Glass: And what Alice Found There, 1893

Für Margarete Henkel und Rolf Herrmann

Den Tod fürchten die am wenigsten, deren Leben den meisten Wert hat.

Immanuel Kant

Table of Contents

Abstract

For a comprehensive understanding of evolutionary processes and for providing reliable prognoses about the future consequences of environmental change, it is essential to reveal the genetic basis underlying adaptive responses. The importance of this goal increases in light of ongoing climate change, which confronts organisms worldwide with new selection pressures and requires rapid evolutionary change to avoid local extinction. Thereby, freshwater ectotherms like daphnids are particularly threatened. Unraveling the genetic basis of local adaptation is complicated by the interplay of forces affecting patterns of genetic divergence among populations. Due to their key position in freshwater communities, cyclic parthenogenetic mode of reproduction and resting propagules (which form biological archives), daphnids are particularly suited for this purpose.

The aim of this thesis was to assess the impact of local thermal selection on the *Daphnia longispina* complex and to reveal the underlying genetic loci. Therefore, I compared genetic differentiation among populations containing *Daphnia galeata*, *Daphnia longispina* and their interspecific hybrids across time, space, and species boundaries. I revealed strongly contrasting patterns of genetic differentiation between selectively neutral and functional candidate gene markers, between the two species, and among samples from different lakes, suggesting (together with a correlation with habitat temperatures) local thermal selection acting on candidate gene *TRY5F* and indicating adaptive introgression. To reveal the candidate genes' impact on fitness, I performed association analyses among data on genotypes and phenotypic traits of *D. galeata* clones from seven populations. The tests revealed a general temperature effect as well as inter-population differences in phenotypic traits and imply a possible contribution of the candidate genes to life-history traits. Finally, utilizing a combined population transcriptomic and reverse ecology approach, I introduced a methodology with a wide range of applications in evolutionary biology and revealed that local thermal selection was probably a minor force in shaping sequence and gene expression divergence among four *D. galeata* populations, but contributed to sequence divergence among two populations. I identified many transcripts possibly under selection or contributing strongly to population divergence, a large amount thereof putatively under local thermal selection, and showed that genetic and gene expression variation is not depleted specifically in temperature-related candidate genes.

In conclusion, I detected signs of local adaptation in the *D. longispina* complex across space, time, and species barriers. Populations and species remained genetically divergent, although increased gene flow possibly contributed, together with genotypes recruited from the resting egg bank, to the maintenance of standing genetic variation. Further work is required to accurately determine the influence of introgression and the effects of candidate genes on individual fitness. While I found no evidence suggesting a response to intense local thermal selection, the high resilience and adaptive potential regarding environmental change I observed suggest positive future prospects for the populations of the *D. longispina* complex. However, overall, due to the continuing environmental degradation, daphnids and other aquatic invertebrates remain vulnerable and threatened.

General introduction

Revealing the molecular basis underlying adaptive responses to environmental changes is a major goal in evolutionary biology. It signifies an essential step towards a comprehensive understanding of evolutionary processes and will allow us to predict the consequences of environmental change for populations and species more accurately (Pardo Diaz *et al.* 2015). In the face of ever increasing anthropogenic influence and disturbance of ecosystems, this goal gains increasing importance. It allows us to infer the degree of ecosystem vulnerability and to derive educated recommendations for future actions (e.g., Whiteley *et al.* 2015). Knowledge of the genetic foundation of traits under selection is therefore not only essential from a conservation biology centered, but also from a social and economic point of view, since the negative impacts of environmental degradation impose an increasing threat on human societies, depriving people of their health, habitats, or food and water resources (IPCC 2014).

Climate change and global warming – vulnerability of freshwater ecosystems

Recent anthropogenic climate change and the accompanying global warming confront populations and species worldwide with new and fast-paced selection pressures (IPCC 2014). This raises concern that many may become extinct and that, as a consequence, the biodiversity of ecosystems will decline (Schneider *et al.* 2007; Bellard *et al.* 2012). According to recent estimates, temperatures are predicted to increase by 2-4°C (depending on the implemented model) within the 21th century (Figure I-1) and extreme weather events are predicted to further increase in frequency. Since temperature is an important selective agent that severely affects organisms through its impact on physical and chemical processes (Pörtner 2001), these developments will be accompanied by changes in a broad spectrum of biotic and abiotic factors (Collins *et al.* 2013; Fischer & Knutti 2015; Mazdiyasni & AghaKouchak 2015).

Inhabitants of freshwater ecosystems are particularly threatened, since they have already severely suffered from anthropogenic stressors such as pollution and increased nutrient influx, as evidenced by the average decline of freshwater species populations included in the Living Planet Index by 50% between 1970 and 2000 (compared to 30% for marine and terrestrial species; Millennium Ecosystem Assessment 2005). Due to rising surface temperatures in temperate lakes, alterations in periods of ice formation in winter (Livingstone & Adrianb 2009; Weyhenmeyer *et al.* 2011), thermocline during summer (Winder & Schindler 2004) and indirect effects such as decreased solubility of gases and acidification in consequence of increased atmospheric CO_2 levels, climate change imposes additional stress upon already disturbed habitats (Dudgeon *et al.* 2006; Settele *et al.* 2014). Species interactions may also play a critical role in climate-related changes in species abundance and local extinctions (Dunn *et al.* 2009; Bellard *et al.* 2012; Vadadi-Fülöp *et al.* 2012).

Additionally, thermal conditions in aquatic habitats are less spatially variable than in terrestrial habitats. Therefore, freshwater habitats offer limited opportunity for organisms to buffer themselves behaviorally from unfavorable thermal conditions, which further aggravates the consequences of global warming on freshwater communities (Gunderson & Stillman 2015).

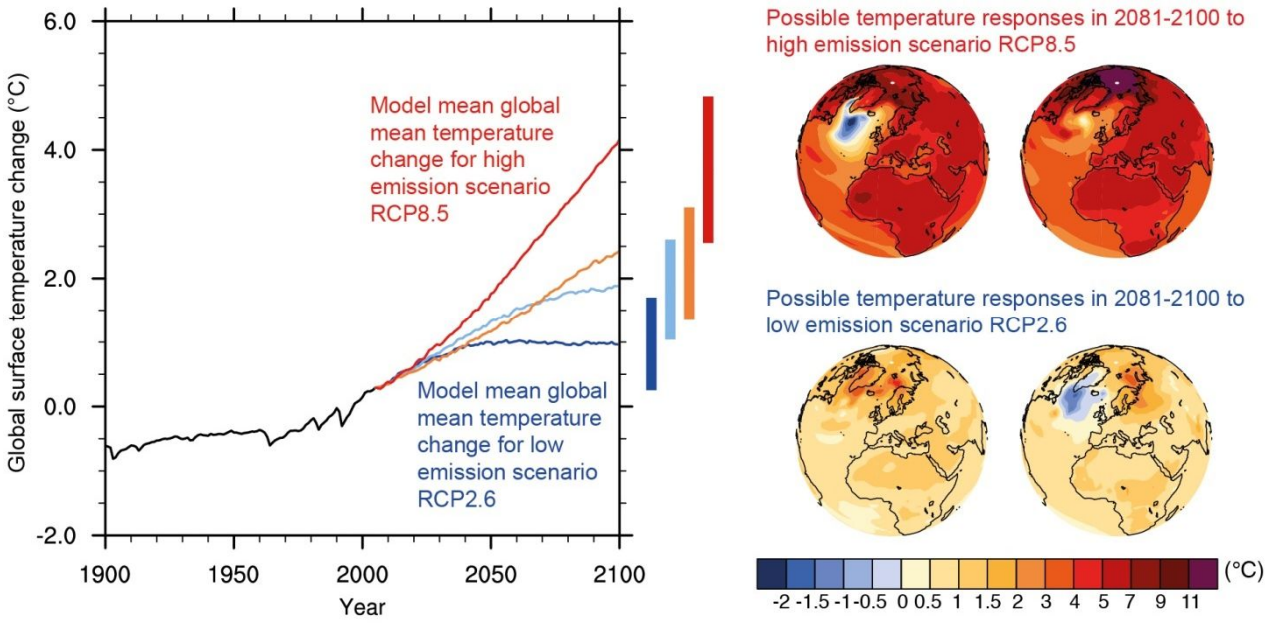

Figure I-1: Global mean temperature change averaged across all Coupled Model Intercomparison Project Phase 5 (CMIP5) models (relative to 1986–2005) for the four Representative Concentration Pathway (RCP) scenarios: RCP2.6 (dark blue), RCP4.5 (light blue), RCP6.0 (orange) and RCP8.5 (red); 32, 42, 25 and 39 models were used respectively for these 4 scenarios. Likely ranges for global temperature change by the end of the 21st century are indicated by vertical bars. Note that these ranges apply to the difference between two 20-year means, 2081–2100 relative to 1986–2005, which accounts for the bars being centred at a smaller value than the end point of the annual trajectories. For the highest (RCP8.5) and lowest (RCP2.6) scenario, illustrative maps of surface temperature change at the end of the 21st century (2081–2100 relative to 1986–2005) are shown for two CMIP5 models. These models are chosen to show a rather broad range of response, but this particular set is not representative of any measure of model response uncertainty. From Collins *et al.* (2013)

The adaptive potential – coping with rapid environmental change

Local extinction of species or populations may be avoided through either acclimatization via phenotypic plasticity or by evolutionary change via genetic adaptation (revieved in, e.g., Dawson *et al.* 2011; Bellard *et al.* 2012; Peñuelas *et al.* 2013). Phenotypic plasticity is the ability of organisms to develop and express particular phenotypic traits in response to environmental cues (Bradshaw 1965). That is, the same genotype may exhibit a variety of, for example, morphological, physiological, behavioral, and phenological characteristics, depending on environmental conditions. These responses may be temporary or permanent throughout a lifespan (Kelly *et al.* 2012b). Evolutionary change via genetic adaptation, by contrast, refers to the emergence of new genetic variation and the change of allele frequencies within a population as a consequence of natural selection (Futuyma 2013). Over the last decades, immense effort was put into documenting

the consequences of climate change, revealing that species respond individualistically rather than cohesively. While several studies indicate rapid evolutionary responses, others show little potential for fast enough adaptation. Cyanobacteria of the genera *Trichodesmium* and *Microcystis*, for example, show rapid adaptation in response to rising CO_2 levels (Hutchins *et al.* 2015; Sandrini *et al.* 2016). *Daphnia magna* show rapid thermal adaptation in response to elevated temperatures (e.g., Salamin *et al.* 2010; Geerts *et al.* 2015) and in *Brassica rapa*, rapidly evolved genetically based changes in flowering time and other phenotypic traits in response to reoccurring drought periods were observed (Franks *et al.* 2016). Examples of species appearing maladapted to changing climate regimes, specifically to elevated temperatures, include lethal effects observed on eggs of the lizard species *Podarcis sicula* and *Calotes versicolor* (Ji *et al.* 2002; Simoniello *et al.* 2016), and low egg-to-adult viability in *Drosophila melanogaster* (Kristensen *et al.* 2015). Lack of an adaptive response over several generations in a selection experiment suggests limited adaptive potential in populations of the copepod *Tigriopus californicus* (Kelly *et al.* 2012a). These examples highlight the heterogeneity of species or populations in their adaptive potential, that is, in their capacity to evolve in response to climate change. The adaptive potential of a population depends mostly on its genetic diversity, i.e., on the amount of standing genetic variation. The strength of selection and the effective population size N_e are also considered important factors, whereby small populations with low N_e are theoretically expected to have reduced adaptive potential: Reduced genetic variation underlying quantitative traits and increased influence of genetic drift in small populations (Lande 1988) are thought to render selection less efficient compared to large populations (e.g., Ellstrand & Elam 1993; Willi *et al.* 2006; Hoffmann & Sgro 2011). However, a recent meta-analysis found no support for the alleged adverse effect of small populations sizes (Wood *et al.* 2016). Phenotypic plasticity contributes little to the adaptive potential, at least with regard to global warming and in ectotherms, where it was shown to be insufficient as protection from temperatures beyond physiological limits (Gunderson & Stillman 2015). Therefore, even if changes in phenotypic traits, such as physiological, morphological or behavioral adjustments, suggest an adaptive response, adaptation may actually be hindered due to low genetic variance. That is because phenotypic variation is often caused by plasticity in phenotypic traits (James *et al.* 1997; Ayrinhac *et al.* 2004; Mason & Taylor 2015) and thus, is not necessarily immediately under selection (Merilä & Hendry 2014). But even with an overall high amount of genetic variation within a population, low levels of genetic variation at a specific locus under selection can prevent an adaptive response (e.g., Hoffmann *et al.* 2003). It is therefore important to not only identify adaptive traits, but also to unveil their underlying genetic basis.

Hybridization and introgression - rapid adaptation though gene flow?

Specifically the standing genetic variation within a population is commonly regarded as the most important factor affecting the adaptive potential in rapidly changing conditions, since it provides readily available genetic diversity for selection to act on (Hedrick 2013). Over extended periods of

time, a rapid rate of environmental change and high selective pressure could, however, outpace the rate of adaptation even in populations that exhibit high genetic diversity. This would decrease the genetic diversity and consequently the adaptive potential of a population. A recent example is the significant loss of genetic diversity in purple sea urchins (*Strongylocentrotus purpuratus*) in response to high levels of carbon dioxide observed in an artificial selection experiment by Lloyd *et al.* (2016). Newly emerging mutations and gene flow, the exchange of genes among populations via migration and subsequent interbreeding, increase the genetic diversity within populations (Hamilton & Miller 2016). Adaptation based on these two processes is usually expected to be slower than adaptation based on standing genetic variation, whereby the pace depends on the mutation rate, the effective population size and the migration rate (Hedrick 2013; Lohr & Haag 2015). Importantly, gene flow may maintain or help recovering the genetic diversity in populations that experienced a loss of standing genetic variation (e.g., Chunco 2014). Gene flow and interspecific hybridization (i.e., interbreeding and gene flow among different species) have therefore gained attention as important evolutionary forces.

Interspecific hybridization is commonly observed in plants, animals, bacteria, and viruses (Arnold 1997). Although regarded as unusual and unimportant in animals for a long time, estimates suggest that approximately 10% of animal species hybridize naturally and many studies show the successful establishment of animal hybrids in nature (Mallet 2005; Abbott *et al.* 2013). Through interspecific hybridization and backcrossing, genetic material can introgress from the gene pool of one species into that of another (Anderson 1949). Environmental change promotes interspecific hybridization and introgression because reproductive barriers such as spatial, temporal or behavioral isolation may become permeable in disturbed habitats. Simultaneously, disturbance might lead to ill-adapted native populations, increasing the chance for successful invasion and colonization. Therefore, elevated levels of hybridization are among the consequences expected from climate change (reviewed in Chunco 2014). Several studies have demonstrated increased levels of hybridization and introgression following both human-mediated and natural environmental change, e.g., in Darwin's finches (*Geospiza* spec.) in response to dietary restrictions following a severe El Niño event (Grant & Grant 1996), in water fleas (*Daphnia* spec.) in response to eutrophication (e.g., Keller *et al.* 2008; Brede *et al.* 2009; Alric *et al.* 2016) and in flying squirrels (*Glaucomys volans*) following a climate change-induced range expansion (Garroway *et al.* 2010).

This combination of elevated stress and hybridization levels may result in the extinction of unique populations or species and thus decrease genetic diversity within a meta-population or species complex (Allendorf *et al.* 2001). Therefore, hybridization and migration have been linked to invasiveness and invasion success in many plant and animal species (e.g., Kolbe *et al.* 2004; Gaskin *et al.* 2009; Turgeon *et al.* 2011; Buhk & Thielsch 2015). Furthermore, gene flow among populations and species may counteract local adaptation by introducing unselected genotypes, dissolving linkage groups and possibly causing outbreeding depression. Consequently, immigration may be detrimental to local populations or biodiversity (Rhymer & Simberloff 1996;

Hendry *et al.* 2001; Seehausen *et al.* 2008; Gómez *et al.* 2015). Yet, interspecific hybridization increases genetic diversity at many loci simultaneously in a single generation and is well known to increase the genetic variability of populations (e.g., Kolbe *et al.* 2004; Facon *et al.* 2008; Lucek *et al.* 2010). Thus, it has the capability to introduce advantageous alleles and increase the adaptive potential of populations confronted with changing conditions. Additionally, as opposed to variation from new mutations, allelic variation introduced by gene flow had already been subjected to selection against deleterious mutations in the donor population. Therefore, hybridization may create novel genotypes and phenotypes at a much faster pace than adaptive change via mutation (Rieseberg *et al.* 2003; Seehausen 2004). Interspecific hybridization can even lead to speciation by transgressive segregation (Rieseberg *et al.* 1999), i.e., the formation of new hybrid taxa that are able to colonize new habitats beyond the range of their parental species. The desert sunflowers *Helianthus anomalus*, *Helianthus deserticola* and *Helianthus paradoxus* are prominent examples of transgressive segregation. All three species originate from hybrids between the two mesic-adapted species *Helianthus annuus* and *Helianthus petiolaris* and occur in either extremely arid conditions or are exposed to high salinity in their habitats (Rieseberg *et al.* 2003). In this example and other recent studies (e.g., Chunco *et al.* 2012; Alric *et al.* 2016), an important role of hybridization for local adaptation was observed, while population or species divergence was maintained even with extensive hybridization and introgression. While certainly often a threat (especially for rare species), interspecific hybridization may contribute to the adaptation of populations confronted with rapid environmental change. Therefore, interspecific hybridization and introgression have been proposed as a possible way to facilitate rapid adaptation in response to anthropogenic disturbance and climate change in many recent studies (Baskett & Gomulkiewicz 2011; Becker *et al.* 2013; Whiteley *et al.* 2015).

Contrasting to plants, interspecific hybrids in animals have long been considered prone to outbreeding depression and less fit than their parental species (Mayr 1963). But since the ancestry of hybrids and their genetic composition can vary profoundly, the fitness of individual genotypes too would be expected to vary substantially (Rieseberg & Carney 1998). Accordingly, individual hybrids actually frequently exhibit hampered fitness compared to their parents (Arnold & Hodges 1995), but there is also an increasing number of studies that document improved fitness in animal hybrids (Lewontin & Birch 1966; Parris 2001; Keller *et al.* 2008; Grant & Grant 2010; Hasselman *et al.* 2014) or through adaptive introgression (Song *et al.* 2011; Pardo-Diaz *et al.* 2012; Lucek *et al.* 2014; Vernot & Akey 2014; Norris *et al.* 2015). However, even in otherwise fit interspecific hybrids, complete or partial loss of sexual reproduction is frequently observed. In facultative parthenogenetic species, which have the ability to reproduce asexually as well as sexually, this can lead to the establishment of clonal lineages. This phenomenon was documented, e.g., in lizards of the *Aspidoscelis cozumela* complex (Manríquez-Morán *et al.* 2014), aphids of the genus *Rhopalosiphum* (Delmotte *et al.* 2003), stick insects of the genus *Bacillus* (Andersen *et al.* 2006), and daphnids of the *Daphnia longispina* complex (Keller *et al.* 2007).

While reduced fertility in hybrids probably prevents speciation via hybridization in many species complexes, it allows for interspecific introgression. Introgression is not equally likely throughout the genome and genome-wide heterogeneity in consequence of interspecific gene flow may persist for a long time even in absence of repeated hybridization events (Stuglik & Babik 2016). Thereby, globally advantageous alleles (i.e., alleles that are equally fit in either genomic background and habitat) are expected to introgress fast. Genetic regions involved in reproductive isolation and local adaptation are, by contrast, expected to impede hybridization and consequently to introgress at a lower rate (reviewed in Harrison & Larson 2014, 2016). Introgression of locally advantageous alleles becomes, however, more likely when local populations become maladapted due to environmental change. Then, migrants from different environmental backgrounds may introduce "preadapted" alleles into native populations and facilitate local adaptation. An example of introgression contributing to local adaptation, although not related to habitat disturbance, are genes regulating wing color patterns in *Heliconius* butterflies. Adaptive, locally mimicry-selected alleles of these genes likely introgressed repeatedly from *Heliconius melpomene* into a clade of *Heliconius cydno* species, as shown in a study by Pardo-Diaz *et al.* (2012).

Because of their potentially severe effects on the genetic diversity within populations and species, positive or negative, and because of their increasing influence under global climate change, processes such as interspecific hybridization and introgression need to be considered in studies on local adaptation and environmental change (Fraisse *et al.* 2016). To evaluate their importance and consequences for populations confronted with environmental change, it is vital to identify the genes underlying adaptive traits and assess their level of introgression.

Assessing adaptive population divergence – approaches and challenges

Sequence-based divergence

Analyses of population divergence are among the most widely used approaches to infer genetic regions potentially underlying traits under local selection in natural populations. Such analyses aim to detect genetic regions that exhibit elevated levels of sequence differentiation or variation in expression levels among populations. At loci that contribute severely to population divergence, the impact of selection is supposed to have exceeded the impact of gene flow and random genetic drift, resulting in elevated sequence differentiation (Beaumont & Balding 2004).

Loci potentially under selection can be detected based on F-statistics (recently reviewed in Hoban *et al.* 2016) or on multivariate, PCA-based analyses (so far rarely used, but see Berthouly-Salazar *et al.* 2016). Other approaches (e.g., the McDonald-Kreitman test, Hudson *et al.* 1987) infer selection by determining the ratios of silent and non-silent polymorphisms (genetic variants with and without an effect on the amino acid sequence of the expressed protein) within populations and substitutions (fixed genetic variants) between populations. Nonetheless, it is challenging to identify adaptive population divergence. Firstly, over short evolutionary time scales, for example, among

populations of the same species, few fixed differences are expected. Secondly, genetic divergence among populations reflects a complex interplay of different factors besides divergent selection. These factors include genetic drift, gene flow, the phylogeographic and demographic histories of populations, the genetic architecture of the selected trait and the strength of selection (recently reviewed in Dittmar *et al.* 2016; Hoban *et al.* 2016). Effects of population structure, demography and incomplete lineage sorting are generally non-randomly distributed across the genome and can severely affect genetic differentiation between populations (Excoffier *et al.* 2009; Lotterhos & Whitlock 2014). Effects of genetic drift may vary among the genome when specific genomic regions contain intrinsic barriers to gene flow (Harrison & Larson 2016). Furthermore, genetic drift is thought to be increased in genomic regions with low recombination rates and consequently locally reduced N_e (e.g., Nosil *et al.* 2009; Pereira *et al.* 2016). As described above, increased genetic drift theoretically results in reduced genetic diversity, weaker impact of selection and increased divergence among populations; this hypothesis is, however, controversial (Wood *et al.* 2016). Apart from bias due to local variation of N_e, genetic drift is a stochastic process and its effects should be randomly distributed across the genome.

The genetic architecture of a selected trait further affects our ability to detect it. Most variation in physiology, behavior, morphology and other complex, fitness-relevant traits is quantitative. Typically, the phenotypic properties of such traits are shaped by multiple quantitative trait loci (QTL, Kearsey 1998). The effect of individual QTL on the phenotype may vary substantially and depends not only on the locus itself, but also on epistasis, i.e., on the interaction between all QTLs contributing to a trait. A large proportion of heritable variation for phenotypes is based on small-effect loci, that is, on genetic regions with small effects on phenotypic variation (Mackay 2001; Rockman 2012), because mutations with small effects on the phenotype are more likely to be beneficial. Large effect mutations are, by contrast, commonly initially deleterious, as suggested by both theoretical and empirical evidence (e.g., Bürger 2000; Barton & Keightley 2002; Orr 2005; Eyre-Walker & Keightley 2007). Yet, methods to infer local adaptation are typically biased towards the detection of large effect loci (with large effects on the phenotype) and thus, might miss important loci because of their small effect sizes (Pardo Diaz *et al.* 2015; Hoban *et al.* 2016). Notably, the severity of this bias, as well as the relative importance of small and large effect loci depend on the adaptive regime: Under slow environmental change, when the population mean trait value is close to the optimum, mostly small effect mutations contribute to local adaptation because large effect mutations are seldomly beneficial. By contrast, in rapidly changing environments, when the population mean trait value is far from the optimum, adaptation is commonly based on large effect mutations (Kopp & Hermisson 2009; Matuszewski *et al.* 2014); particularly in presence of gene flow (Griswold 2006; Yeaman & Whitlock 2011). A large mutation can rapidly reduce the large difference between the optimal and the actual phenotype arising through fast environmental change. Thus, rapid fixation by a selective sweep is more likely if the selected trait is based on few large-effect loci (Kopp & Hermisson 2007).

Several F_{ST}-based methods test for selection while accounting for random effects by estimating the level of differentiation expected under neutrality (i.e., attributable to selective neutral demographic and genetic parameters such as population size, migration, and mutation rates and recombination) and identifying outlier loci exceeding this neutral level of population differentiation (Antao *et al.* 2008; Foll & Gaggiotti 2008). Because such tests do not account for the non-random factors influencing population divergence described above, and cannot detect selection at loci exhibiting a divergence level as low as expected under neutrality or lower, these tests are prone to false positives as well as to false negatives. Therefore, tests that also incorporate ecologically relevant (Rellstab *et al.* 2015) or phylogenetic information (Bonhomme *et al.* 2010) were introduced. These methods account for several of the described factors but require data that are not always available, like outgroup genomes and detailed environmental data. In spite of the drawbacks, outlier tests have been used to find candidate genes under selection in a great amount of studies on trait variance and adaptation over the last decade, for example, in insects (Chavez-Galarza *et al.* 2013; de Jong *et al.* 2013; Sedghifar *et al.* 2016), mussels (Luttikhuizen *et al.* 2012; Gosset & Bierne 2013; Eierman & Hare 2016; Fraisse *et al.* 2016), fishes (Raeymaekers *et al.* 2007; Shikano *et al.* 2010; Vera *et al.* 2016), amphibians (Zielinski *et al.* 2014; Yang *et al.* 2016b), birds (Parchman *et al.* 2013; Mason & Taylor 2015; Szulkin *et al.* 2016), and mammals (Pariset *et al.* 2009; Henry & Russello 2013; Roffler *et al.* 2016; Yang *et al.* 2016a). Candidate gene approaches can be integrated into population genetic analysis to assess the impact of specific environmental factors on local adaptation (Shimada *et al.* 2011; Berdan *et al.* 2015; Berthouly-Salazar *et al.* 2016). Thereby, the combination of different detection methods may help to reduce the number of false positive results. Nonetheless, it is essential to demonstrate change over time and furthermore, to link identified outlier loci to a population-specific fitness advantage to evidence a contribution of such loci to local adaptation (Hansen *et al.* 2012), e.g., by follow-up genotype-phenotype association studies (e.g., Du *et al.* 2014; Chaves *et al.* 2016).

Expression-based population divergence

Phenotypic variation within and between species, including differences in phenotypic plasticity, often result from heritable variation in expression levels (Wray *et al.* 2003; Romero *et al.* 2012; Pai *et al.* 2015). This variation can be assumed to be ultimately caused by genetic diversity in loci associated with regulatory mechanisms, which interfere with gene expression during or after transcription (Pai *et al.* 2015). Such regulatory mechanisms include the transcription factor binding affinity, which influences the transcription rate of a gene; histone modifications and DNA methylation, which both influence gene accessibility for transcription; and gene duplications, which increase the amount of transcribed mRNA. Post-transcriptionally, RNA degradation and alternative splicing regulate the amount and composition of available mRNA in a cell (Harrison *et al.* 2012). The regulatory elements underlying these mechanisms may be located in close proximity to the gene they regulate (*cis*-regulatory elements), e.g., within introns, 5' and 3' UTRs, or tens of

kilobases up- or downstream, but may also be located elsewhere in the genome, i.e., far away from the regulated gene (*trans*-regulatory elements; Wray, 2007). Regulatory elements frequently affect the expression of several genes and some affect transcript abundances of hundreds to thousands of genes. This phenomenon of one locus affecting multiple other loci (or traits) is referred to as pleiotropy (Paaby & Rockman 2013).

Since selection acts on the expression as well as on the sequence level, gene expression patterns often play an important role in the evolution of complex traits and contribute to species-specific adaptations (Gilad *et al.* 2006; e.g., Whitehead & Crawford 2006; revieved in Wray 2007). Adaptive changes in expression levels and DNA sequences possibly represent alternative options for responses to selection pressure, which may be specifically relevant if either of these options is constrained due to negative pleiotropic effects (Shapiro *et al.* 2004; Harrison *et al.* 2012). Differentiation of gene expression is probably specifically relevant in early stages of divergence among populations (Pavey *et al.* 2010; Brawand *et al.* 2011). Loci exhibiting divergent expression levels among populations may therefore signify divergent selection for regulatory differences (e.g., Gilad *et al.* 2006). However, elevated levels of expression level variation might alternatively reflect a lineage-specific relaxation of evolutionary constraint instead of selection (Romero *et al.* 2012).

Transcriptome analyses allow for the identification of genetic regions that exhibit divergence in expression levels, including pleiotropically regulated genes. Using RNA sequencing (RNA-seq), transcriptome-wide expression level variation can be assessed simultaneously with coding-sequence diversity (Wang *et al.* 2009; Todd *et al.* 2016). The transcriptome encompasses all RNA molecules expressed in a cell, a population of cells or an organism (Velculescu *et al.* 1997). In contrast to the genome, the transcriptome changes dynamically in response to environmental or intrinsic factors (Richards *et al.* 2009), influenced, e.g., by parasites (Feldmeyer *et al.* 2016), abiotic factors, age (Landis *et al.* 2012), sex (Liu *et al.* 2015) or circadian rhythm (Nolte & Staiger 2015). Accordingly, most transcriptome studies on the genetic basis of particular phenotypic traits compare responses to various treatments within a single group of specimen or between tolerant and sensitive specimen, but not among populations. Examples of such studies are analyses of differential expression among four fitness-related *D. melanogaster* phenotypes exposed to various treatments (including different physical and chemical stressors as well as feeding and social crowding conditions) by Zhou *et al.* (2012) and the study of Li *et al.* (2016) on transcriptomic responses of freezing tolerant and sensitive lines of maize (*Zea mays*) to freezing treatment. Such studies provide candidate genes important in coping with various environmental conditions, but reveal little information about local adaptation or the populations' adaptive potential. The number of population-transcriptomic publications is, however, increasing. Von Heckel *et al.* (2016), for example, identified potential targets of cold adaptation in *D. melanogaster* by comparing the response to cold shock and control conditions in tropical and cold-temperate populations. They suggested that canalization of gene expression is responsible for the increased cold tolerance

observed in the cold-temperate population. Divergent expression levels were also observed in response to a simulated heat wave scenario between the two co-occurring seagrass species *Zostera marina* (less stress-tolerant) and *Nanozostera noltii* (more stress-tolerant). A comparison of northern and southern populations of both species revealed population-specific responses only in *N. noltii* (Franssen *et al.* 2014). In a follow-up study on two North-South pairs of *Z. marina* populations, however, Jueterbock *et al.* (2016) showed faster recovery from heat-stress and parallel transcriptomic differentiation in the southern populations.

Daphnia, a model system for ecology and evolution

Daphnia (Crustacea: Cladocera: Anomopoda), small planktonic crustaceans, which feed mainly on algae, are particularly suited for research on the molecular basis underlying adaptive responses to environmental change (Lampert 2011). *Daphnia* species hold a key position in the aquatic food web, since they dominate the zooplankton community in many temperate freshwater bodies and serve as links between primary producers and secondary consumers, such as zooplanktivorous fish and insects (e.g., Sterner *et al.* 1992; Jeyasingh *et al.* 2011). Another particularly advantageous characteristic of *Daphnia* is their cyclic parthenogenetic mode of reproduction, which is found in the majority of species within the genus (Hebert 1978). Under favorable environmental conditions, daphnids reproduce asexually, i.e., produce clonal offspring via parthenogenesis. This property allows for the establishment of clonal lineages in the laboratory and hence for replicated experimental studies on genotype-specific phenotypic responses to various treatments. Harsh environmental conditions typically induce the sexual mode of reproduction, which involves the production of ephippia (Banta *et al.* 1939; Figure I-2 A and B). Ephippia are thick chitinous shells that contain two long-lived, dormant propagules and accumulate in the lake sediment (Hebert 1978), forming temporally stratified biological archives (Figure I-2 C). This process conserves the genomic histories of *Daphnia* populations and their communities, along with information on environmental parameters entrapped in the sediment soils (e.g., eutrophication, oligotrophication, and pollution levels; Kerfoot 1974). Therefore, by dating of sediment layers, responses to documented or inferred environmental change can be traced trough time (reviewed in Orsini *et al.* 2013), a necessity for the reconstruction of evolutionary processes that contributed to present-day population genomic structures (Hansen *et al.* 2012). Furthermore, the sturdiness of ephippia and longevity of propagules make it possible to resurrect past populations in the laboratory and to infer their phenotypic responses under various treatments (resurrection ecology, Kerfoot *et al.* 1999). This property was used prominently to demonstrate Red Queen dynamics by resurrecting parasites along with *D. magna* from sediment samples and comparing responses to contemporaneous, older and younger parasites (Decaestecker *et al.* 2007). Another example is the assessment of evolutionary responses to cultural eutrophication in *Daphnia pulicaria*, where specimen potentially as old as 700 years were resurrected and shown to exhibit higher growth efficiency under low phosphorus diets compared to modern genotypes (Frisch *et al.* 2014). This combination of

advantages, together with the vast resources of available information – *Daphnia* is not only one of the most extensively studied subjects in ecology (Ebert 2005), but has also two fully sequenced genomes by now (*Daphnia pulex* genome, Colbourne *et al.* 2011, Ye *et al.* 2017) - make *Daphnia* a prominent model system in aquatic ecology and evolution (Lampert 2011).

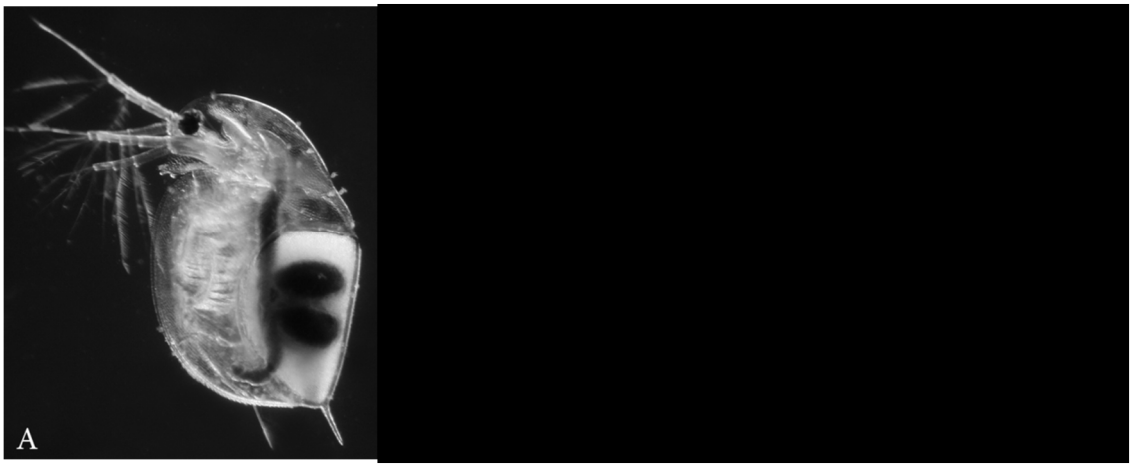

Figure I-2: Sampling of *Daphnia* ephippia from sediment: A: sexual *D. longispina* female carrying an ephippium; B: photographs and schematic drawings of ephippia of *D. pulex*, *Daphnia obtusa* and *D. longispina* (from top to bottom); C: stratified lacustrine sediment cores. A by D. Ebert; B and C modified from Mergeay *et al.* (2005) and Orsini *et al.* (2013)

Local adaptation and population structure in Daphnia

In organisms with high dispersal capabilities, such as daphnids, levels of gene flow among populations are expect to be high (Mayr 1942). Daphnids disperse mainly through transportation of ephippia by wind (Cáceres & Soluk 2002) or waterbirds (Figuerola *et al.* 2005). Nonetheless, a large variety of studies, including analyses of allozyme variation (e.g., Spitze 1993; Vanoverbeke & De Meester 1997; Haag *et al.* 2005), microsatellite DNA (e.g., Thielsch *et al.* 2009; Allen *et al.* 2010; Griebel *et al.* 2016), and genome-wide markers (Orsini *et al.* 2012), revealed substantial genetic differentiation among *Daphnia* populations, indicating low levels of gene flow. Although intuitively paradoxical, such substantial divergence is common among populations of aquatic species with high dispersal capabilities via resting propagules and high reproduction rates (De Meester *et al.* 2002). According to the Monopolization Hypothesis (De Meester *et al.* 2002; De Meester *et al.* 2016), this observation is probably explained by a combination of founder and priority effects and local adaptation: The term "founder effects" refers to the establishment of a population based on few individuals or propagules, which is, due to genetic drift, distinctively different from the population from which the founder genotypes originated (Mayr 1942; Wright 1942). Priority effects, i.e., the impact of arrival order on a community (Shulman *et al.* 1983) or population (Van Gremberghe *et al.* 2009), reinforce founder effects. Their influence is particularly severe, if reproduction rates are high and the carrying capacity is reached quickly, which gives early colonists

a numerical advantage over later arriving migrants. Consequently, these migrants have a low probability of establishment success and of contributing significantly to the gene pool. This probability may be further reduced by local adaptation of the resident population and by the establishment of a large egg bank of locally adapted propagules. These can repopulate the habitat immediately when the carrying capacity increases and can likely outcompete non-adapted competitors. Particularly in species with a cyclic parthenogenetic mode of reproduction, like *Daphnia* species, the combination of clonal selection (allowing for a fast spread of fit genotypes) and sexual recombination (providing continuous supply of new genotypes) may effectively minimize gene flow and elevate genetic differentiation among populations (De Meester *et al.* 2002). Differentiation among *Daphnia* populations is further influenced by the rate of interspecific hybridization. Thereby, large proportions of hybrids may additionally increase divergence among populations, because hybrids often have limited success in sexual reproduction, produce few ephippia and thus, have low dispersal capabilities (e.g., Yin *et al.* 2014; Griebel *et al.* 2016).

Divergence among *Daphnia* populations is, however, not limited to DNA sequence differentiation, but has also been shown in numerous phenotypic traits. Phenotypic divergence among populations that suggests local adaptation has been shown, e.g., in resistance to toxic cyanobacteria (Sarnelle & Wilson 2005; Jiang *et al.* 2015), life history responses to several stressors (Boersma *et al.* 1999) and sex induction (Roulin *et al.* 2013). Several studies took advantage of the dormant propagules archived in pond sediment, compared populations across different points in their past and thereby demonstrated phenotypic change over time. Examples are changes in migration behavior that correlate with changes in fish predation pressure (Michels *et al.* 2007) and increased toxin resistance that corresponds to increased levels of toxic cyanobacteria (Hairston *et al.* 1999). Evidence for adaptive change over time beyond correlations was produced, e.g., for phototaxis, a trait important in predator-avoidance. Combining a resurrection ecology and a F_{ST}-Q_{ST} approach, Cousyn *et al.* (2001) compared population differentiation at putatively selectively neutral genetic markers and for an adaptive quantitative trait. Thereby, they demonstrated increased plasticity in phototactic behavior in response to fish kairomones due to natural selection by planktivorous fish.

However, despite growing interest and efforts, verifying local adaptation of phenotypic divergent traits and particularly pinpointing the genetic basis of locally adapted traits in *Daphnia* has been mostly inconclusive so far. Usually, either the genetic basis of an adaptive trait remains unknown (e.g., Cousyn *et al.* 2001; Altermatt *et al.* 2007; Yampolsky *et al.* 2014a) or the loci underlying potentially selected traits are known, but the evidence for adaptation remains insufficient (Williams *et al.* 2012; Yampolsky *et al.* 2014b).

The influence of temperature on Daphnia

As aquatic ectotherms, daphnids are particularly threatened by increasing temperatures (Gunderson & Stillman 2015). Although ectotherms possess a range of physiological and behavioral responses to cope with the impact of temperature, they are likely to experience

temperatures beyond their physiological limits as a result of climate change (Gibert & De Jong 2001; Gunderson & Stillman 2015). Therefore, *Daphnia* populations must adapt rapidly to escape local and possibly global extinction in the face of ongoing global warming. Examples of phenotypic traits affected by temperature in *Daphnia* include life-history processes, metabolism, and gene expression (reviewed in Wojtal-Frankiewicz 2012). Phenotypic plastic responses to increased temperatures include, e.g., increased locomotion and filtering rates and are associated with elevated respiration (Yurista 1999). Because of their high energetic cost, such activities cannot always be maintained over an extended period and furthermore, respiratory processes are adversely affected by temperatures beyond a certain threshold. Consequently, high temperatures are often lethal for daphnids (Foran 1986; Zeis *et al.* 2013; Schumpert *et al.* 2014) or reduce the proportion of energy available for growth and reproduction (Yurista 1999). This may result in a decline of indigenous species (Lennon *et al.* 2001). In addition to temperature increase, the effect of climate change on community composition and species interactions adds further selective pressure. Particularly the alteration of the food web structure due to rapid adaptation of green algae and toxic cyanobacteria to rising temperatures, CO_2 levels, and phosphorus content (Paerl *et al.* 2011; Padfield *et al.* 2016; Sandrini *et al.* 2016) will have major effects on *Daphnia* populations.

Thus far, *Daphnia* populations seem to be capable of responding to these threads by adaptive change, as documented by several studies: Adaptive change in response to temperature increase, namely of genotype composition (Van Doorslaer *et al.* 2009b), life-history traits (Van Doorslaer *et al.* 2010; Henning-Lucass *et al.* 2016) and temperature tolerance (Geerts *et al.* 2015), has been shown in selection experiments as well as over time in natural populations. Recent studies suggest local thermal adaptation in *D. pulex* and *D. magna* (Williams *et al.* 2012; Yampolsky *et al.* 2014a) and enhanced tolerance to elevated temperatures in *D. magna* from warmer climates due to adaptive plasticity in hemoglobin expression (Yampolsky *et al.* 2014a). Yampolsky *et al.* (2014b) revealed candidate genes possibly underlying thermally selected traits by focusing on genome-wide expression level variation among heat-tolerant and heat-sensitive *D. magna* clones in response to optimal and severely increased temperatures. However, the contribution of these genes to phenotypic and genetic population divergence and to local adaptation remains unclear, because of the high inter-clonal variation commonly found within *Daphnia* populations. Inter-clonal variation may exceed variation between populations, which includes variation observed in response to temperature (Mitchell & Lampert 2000). Differences in performance among populations in intermediate temperatures have been shown in metabolic rates (Chopelet *et al.* 2008) and might play a major role under natural conditions (Mitchell & Lampert 2000). Several gene families, specifically hemoglobins (Paul *et al.* 2004b) and proteases (Schwerin *et al.* 2009) seem to be relevant for coping with different thermal conditions in *Daphnia*, mainly because of elevated oxygen and energy demand in elevated temperatures. Notably, proteases are not only often involved in stress responses (Asselman *et al.* 2015) but also relevant because of their varying sensitivity to cyanobacteria toxins (e.g., Schwarzenberger *et al.* 2012; De Coninck *et al.* 2014).

Williams *et al.* (2012) demonstrated that differences in expression of temperature-relevant genes may signify adaptive canalized expression. In their study on *D. pulex*, heat tolerant clones possessed an increased level of hemoglobin expression (enabling their enhanced tolerance) compared to heat-sensitive clones, even at non-stressful temperatures. However, the genes involved in thermal adaptation apart from acute stress responses, and the relative importance of temperature-relevant genes in local adaptation remain to be revealed.

The Daphnia longispina *species complex*

Daphnia species frequently form interspecific hybrids and are accordingly grouped into species-hybrid complexes (Schwenk & Spaak 1995). A species complex exhibiting a wide distribution and particularly common interspecific hybridization throughout Central Europe is the *D. longispina* complex (Schwenk 1993; Gießler *et al.* 1999; Taylor *et al.* 2005; Ishida *et al.* 2011). It includes, amongst several cryptic lineages, at least six different species, including *Daphnia galeata* and *Daphnia longispina* (Figure I-3; taxonomy revised in Petrusek *et al.* 2008a).

Figure I-3: *D. galeata* and *D. longispina*. A: *D. galeata*: 1. Ex-ephippio female, 2. Rostrum of 1; 3. Summer female, 4. Rostrum of 3; B: *D. longispina* (formerly *D. hyalina*): 1. Female, 2. Head of ex-ephippio female, 3. Summer female. Modified from Flößner and Kraus (1986)

D. galeata typically inhabits large lakes and dominates in eutrophic conditions, while *D. longispina* also occurs in smaller ponds and dominates in oligotrophic lakes (Flößner 1972; Flößner & Kraus 1986; Brede *et al.* 2009; Thielsch *et al.* 2009). According to Keller *et al.* (2008), higher temperatures possibly promote *D. galeata* dominance. Despite ongoing interspecific gene flow, the species show clear genetic differentiation (Petrusek *et al.* 2008a; Thielsch *et al.* 2009). Interspecific hybrids occur syntopically with their parental species in a patchy, but geographically widespread distribution (Schwenk *et al.* 2000). Thereby, several studies observed predominantly unidirectional

hybridization of *D. galeata* females and *D. longispina* males (Schwenk 1993; Brede *et al.* 2009; Griebel *et al.* 2016), but there are also reports indicating bidirectional hybridization (Taylor & Hebert 1993; Alric *et al.* 2016).

Interspecific hybrids of the *D. longispina* complex often show reduced success in sexual reproduction and thus, contribute little to the resting egg bank (Keller *et al.* 2007; Griebel *et al.* 2016), but they often become successfully established in nature. They compete successfully with their parental species during the parthenogenetic phase of the life cycle (Weider & Wolf 1991; Boersma & Vijverberg 1994; Seidendorf *et al.* 2007) and even exceed their fitness (Ebert *et al.* 2002; Griebel *et al.* 2015; Lohr & Haag 2015). Griebel *et al.* (2015) observed, for example, a higher survival probability under simulated winter conditions in hybrids from three different lakes compared to parthenogenetic females of the parental species. Despite a much higher rate of clonal reproduction and producing fewer sexual diapause eggs compared to parental species (Keller *et al.* 2007; Yin *et al.* 2010), interspecific hybrids may further backcross with parental species (Jankowski & Straile 2004; Keller & Spaak 2004). This can lead to substantial introgression among taxa (Petrusek *et al.* 2008a).

Severe environmental fluctuations (associated with a high mortality) influence the relative fitness of *Daphnia* hybrids, e.g., in harsh winters (Griebel *et al.* 2016), and may reduce hybrid diversity over time because interspecific hybrids contribute little to the resting egg banks and therefore to the repopulation of habitats. Less severe environmental change often promotes interspecific hybridization events, as demonstrated by Jankowski and Straile (2003), Brede *et al.* (2009), Rellstab *et al.* (2011) and Alric *et al.* (2016). Through the analysis of resting egg banks, these studies revealed dramatic changes in the taxon composition of the *Daphnia* community over time in several European lakes in response to eutrophication, associated with hybridization and subsequent nuclear as well as mitochondrial introgression between *D. galeata* and *D. longispina*. They suggest that adaptive introgression allows for combining locally adapted, indigenous genotypes with preadapted, immigrating genotypes, which might contribute to the semipermeable nature of the reproductive barrier between these hybridizing species. Hence, interspecific hybridization and introgression have been proposed as possible mechanisms of rapid evolutionary change in natural *Daphnia* populations (Brede *et al.* 2009; Griebel *et al.* 2016). However, despite the evidence for abundant hybridization events and the potential importance of interspecific introgression on adaptation of *Daphnia* populations, little is known about the involved genetic regions.

Thesis outline

Aims of this study

The aim of this thesis is to assess the impact of thermal selection on populations of the *D. longispina* complex and to reveal genetic loci underlying local thermal adaptation. Therefore, I address the following overarching questions:

1. Are natural populations of the *D. longispina* species complex adapted to local thermal regimes?
2. Are interspecific hybridization and introgression contributing to local thermal adaptation?
3. What conclusions can be drawn regarding the fate of the *D. longispina* species complex in light of advancing climate change?

Chapter overview

Specifically, I focus in **Chapter 1** on three putatively fitness-relevant and ecologically important candidate genes and compare them to presumably neutral genetic markers. To that aim, I genotyped resting eggs obtained via sampling of lake sediments. This allowed for comparisons among populations comprised of *D. galeata*, *D. longispina*, and their interspecific hybrids along a latitudinal cline across Europe and, in two populations, over time. Thereby, I revealed strongly contrasting patterns of genetic differentiation among the different types of genetic markers, which provides insight regarding the genetic regions affected by local selection and differential introgression among species. Drawbacks of this candidate gene approach are the lack of information regarding the detected allelic variants. Both their historic distribution within and among populations, and their impact on organismal fitness need to be examined further.

Therefore, I focus in **Chapter 2** on the relationship between potentially adaptive alleles and fitness by inferring genotype-phenotype relationships at the candidate gene loci examined in **Chapter 1.** I genotyped experimental *D. galeata* clones from several European populations with information on phenotypic traits in response to different temperature regimes. This approach allowed me to circumvent the drawbacks inherent to monitoring either phenotypic shifts or molecular markers separately; namely the lack of information about the role of plasticity in trait shifts and of molecular markers in individual fitness. The association analyses allow me to draw conclusions about the extent of the candidate genes' contribution to phenotypic traits and offer insights into the role of phenotypic plasticity in response to thermal variation.

Finally, I utilize a combination of a population transcriptomic and a reverse ecology approach, described in **Chapter 3**, to assess the role of thermal selection in shaping population divergence in *D. galeata*. To that aim, I screened the literature on arthropods for genes putatively involved in thermal adaptation and assembled a comprehensive set of temperature-relevant candidate genes. Through the comparison of transcriptomes among four populations of *D. galeata*, I was able to

identify transcripts possibly under selection or contributing substantially to population divergence on the levels of both coding sequences and gene expression. Analyses of the proportions of temperature-relevant candidate genes among such transcripts allowed me to elucidate the influence of local thermal regimes on divergence among *Daphnia* populations, to identify a large amount of transcripts that possibly contribute to local thermal adaptation in *D. galeata* and to compare the influence of thermal selection on the coding sequence and on the gene expression level.

Chapter 1

Natural selection in *Daphnia* populations across time, space, and species boundaries

Maike Herrmann, Mathilde Cordellier, Pia Kreuzer, and Klaus Schwenk

Abstract

Populations facing changing environmental conditions can evade local extinction by exhibiting phenotypic plasticity or by local (genetic) adaptation. Interspecific hybridization has been proven to be a potentially important contributor to adaptation and processes such as interspecific hybridization and needs to be considered in studies on local adaptation. In *Daphnia* species complexes, frequent interspecific hybridization events and the occurrence of introgression among species are well documented, but little is known about the gene regions involved. Therefore, we studied genetic differentiation in a hybrid complex using neutral (mtDNA and microsatellite) loci as well as three ecologically relevant candidate genes that exhibit proteomic responses to variation in temperature and possibly contribute to increased tolerance to cyanobacterial toxins.

Network analyses revealed strongly contrasting patterns of genetic differentiation among neutral, phylogenetically informative loci, and ecologically relevant candidate gene loci. Different outlier tests showed a strong signal for directional selection at locus *TF* (trypsin) and analyses of molecular variance revealed higher genetic differentiation among sampling localities than among hybridizing species at this locus. Analyses of genetic differentiation among populations from different geographic areas as well as from different time periods revealed no pattern of isolation by distance and either none or moderate differentiation over time.

Overall, the contrasting patterns of genetic divergence among marker types and taxa of the *Daphnia longispina* species complex most likely result from introgression. We provided evidence for selection driving population differentiation at candidate locus *TF*, which demonstrates that introgression is not restricted to selectively neutral regions of the genome in taxa of the *D. longispina* species complex. Our results support the hypothesis that differential introgression of preadapted alleles might enable or facilitate local adaptation to current environmental changes in this species complex. Thus, they enhance our understanding of the interplay of evolutionary forces affecting local adaptation processes as well as the permeability of reproduction barriers among species.

Submitted to *BMC Evolutionary Biology*

Introduction

Understanding the micro-evolutionary basis of adaptive responses to environmental changes is a major goal in evolutionary biology. Lately, a number of studies aimed at identifying genetic regions underlying adaptation, many of them by comparing adaptive differentiation between several populations of a species (reviewed in Barrett & Hoekstra 2011) and with a strong emphasis on global change (see, e.g., Merilä & Hendry 2014 and reviews therein). In this context, local adaptation is usually studied without considering interspecific hybridization and immigration is still often viewed as solely detrimental to local populations or species as it puts them at risk of displacement (e.g., Van Doorslaer *et al.* 2009b; Andrew *et al.* 2012). Hybridization and differential introgression among closely related species are, however, now recognized as common in both plants and animals and have gained acceptance as potentially important and creative forces in evolution (reviewed in Harrison & Larson 2014, 2016).

Generally, hybridization introduces genetic variation at many loci in a single generation and the newly introduced allelic variation has been subjected to selection against deleterious mutations in the genetic background of the donor population (Rieseberg *et al.* 2003). Since the innovative study of Rieseberg *et al.* (2003) on ancient *Helianthus* hybrids, improved fitness through adaptive introgression was shown in a number of studies and species (e.g., Whitney *et al.* 2006; Song *et al.* 2011; Pardo-Diaz *et al.* 2012). However, in current literature, only globally advantageous alleles, i.e., alleles that are equally fit in either genomic background and habitat are expected to introgress fast. Introgression of genes involved in reproductive isolation and local adaptation, i.e., loci underlying locally divergent selection is, by contrast, expected to be hampered (reviewed in Harrison & Larson 2014, 2016). Natural selection likely prevents migrants from displaying higher fitness over residing, already locally well adapted genotypes. However, the role of adaptive introgression in locally mimicry-selected wing color patterns in *Heliconius* butterflies (Pardo-Diaz *et al.* 2012) demonstrates that there are exceptions to this rule.

Introgression of gene regions involved in local adaptation becomes more likely especially when populations are confronted with environmental disturbances. Several studies have demonstrated increased levels of hybridization and introgression following both human-mediated and natural environmental change (e.g., Grant & Grant 1996; Brede *et al.* 2009; Ishida *et al.* 2011). In disturbed habitats, hybridization has been suggested as a way to overcome the loss of genetic diversity in a population (Brede *et al.* 2009; Whiteley *et al.* 2015). Immigrants from a preadapted population might not outcompete the indigenous population, but rather provide a fast and effective way for the indigenous recipient population to regain fitness.

The waterflea species *Daphnia galeata* and *Daphnia longispina*, belonging to the *Daphnia longispina* complex, are particularly suitable for studying interspecific hybridization and differential introgression (Schwenk 1993). Despite the clear genetic differentiation among species (Petrusek *et al.* 2008a; Thielsch *et al.* 2009), hybridization within the species complex is common

(Schwenk 1993; Gießler *et al.* 1999; Taylor *et al.* 2005; Ishida & Taylor 2007b; Dlouha *et al.* 2010; Ishida *et al.* 2011; Thielsch *et al.* 2012). Across Europe, interspecific hybrids occur syntopically with their parental species in a patchy, but geographically widespread distribution (Schwenk *et al.* 2000). Interspecific hybrids thereby often reach high abundances (Schwenk & Spaak 1995) and, importantly, frequently exhibit increased fitness under particular ecological conditions compared to parental species (Ebert *et al.* 2002; Griebel *et al.* 2015; Lohr & Haag 2015). Through the analysis of resting egg banks in Lake Constance and Greifensee, introgression in *Daphnia* has been observed over time and in response to environmental change (Brede *et al.* 2009). This study suggests that adaptive introgression allows combining locally adapted indigenous with preadapted, immigrating genotypes, which seems a plausible mechanism for preventing the build-up of a more pronounced reproductive barrier between hybridizing species.

Despite the evidence for abundant hybridization events and interspecific introgression in *Daphnia*, hardly anything is known about the nature of the gene regions involved. We therefore chose to analyze a combination of genetic markers differentially affected by hybridization and selection. Additionally to mitochondrial DNA and neutral microsatellite markers, three candidate genes coding for proteins potentially involved in developmental, behavioral, and physiological responses to changing temperatures and cyanobacterial toxin were analyzed. These candidate genes are likely of ecological importance and relevant to individual fitness, as evidenced by experimental results (Agrawal *et al.* 2005; Schwerin *et al.* 2009; Asselman *et al.* 2012; De Coninck *et al.* 2014; Asselman *et al.* 2015). Schwerin *et al.* (2009) showed that the proteins coded by these genes, were shown to be significantly differentially expressed in *Daphnia pulex* at 10 °C and 20 °C. The gene *MHC-1* encodes for the muscle myosin heavy chain, a part of the motor protein myosin, which is involved in locomotion and filtration. The gene *ERNA* encodes for an enzyme carrying the EndoU/XendoU domain characteristic for a family of endoribonucleases involved in ribosome biogenesis (Renzi *et al.* 2006); thus directly affecting translation and potentially influencing development and physical responses to environmental changes. Last, the gene *TRY5F* encodes for a trypsin, a secretory enzyme involved in protein digestion and possibly important for resistance to cyanobacterial toxins (Agrawal *et al.* 2005; Asselman *et al.* 2012; De Coninck *et al.* 2014; Asselman *et al.* 2015).

By focusing on ecologically important and fitness-relevant candidate genes we applied a hypothesis-driven strategy (comparable to Eveno *et al.* 2008; Hamilton *et al.* 2013a). If interspecific hybridization contributed to genetic variation at candidate loci, we expect these loci to show a lack of genetic differentiation among the hybridizing species *D. galeata* and *D. longispina*. Should differential introgression have occurred, we additionally expect high levels of genetic differentiation among species at putatively neutral markers. If the candidate genes are under selection, we expect them to exhibit distinct patterns of genetic differentiation among sampling localities in comparison with putatively neutral microsatellite markers.

Sampling sites are distributed along a latitudinal cline across Europe, spanning mean annual temperatures between 8.8 °C and 15.9 °C. We obtained propagules of *D. galeata*, *D. longispina* and their interspecific hybrids from layered sediments and genotyped them using the putatively neutral and non-neutral markers described above to infer genetic variation among populations. Within two populations, we also inferred genetic variation over time (temperatures increased between 0.4 °C and 1.5 °C). Although experimental tests commonly involve temperature differences of 4 °C (Van Doorslaer *et al.* 2009b; Van Doorslaer *et al.* 2010; Geerts *et al.* 2015) to 10 °C (Schwerin *et al.* 2009), recent climate change apparently adds sufficient selective pressure to evoke adaptational change. Recent studies on *D. galeata* and *Daphnia magna* clones from different time periods suggest temperature-driven micro-evolution over time in response to differences in mean annual temperatures of about 1.2 °C (Geerts *et al.* 2015; Henning-Lucass *et al.* 2016).

We subsequently used the estimated levels of genetic differentiation among these natural *Daphnia* populations to infer the footprints of introgression and selection across time and space for the different genetic markers. Finally, we tested for isolation by distance and environment to identify the potential origin of genetic differentiation among populations.

Results

Species affiliation based on neutral and candidate loci

Genotype data of 11 microsatellite loci were analyzed for species affiliation of *D. longispina* and *D. galeata* individuals and identification of interspecific hybrids in each of the sixteen pools of individuals. Throughout this paper, "population" refers to a pool of individuals from a lake (Figure C1-1), whereas individual pools from different layers within the same lake will be addressed as "subpopulations". The number of individuals analyzed is given in detail in Table C1-1. A covariance-standardized principal coordinate analysis (PCoA) was then conducted to find patterns of neutral genetic differentiation and to visualize the distribution of genetic variation within and among taxa (see Additional file C1-2: Table S1 for taxon abundances). This revealed a pattern of two distinct species clusters with interspecific hybrids distributed between them at neutral nuclear loci (PCoA of multilocus genotypes; Figure C1-2B).

Subsequently, we used sequences obtained for the mitochondrial rDNA locus *12S* and the three candidate loci *ERNA* (*EA*), *TRY5F* (*TF*), and *MHC-1* (*M*) to analyze intra- and interspecific relationships among haplotypes. Therefore, *12S* rDNA and candidate loci haplotype networks were generated. In *12S* rDNA haplotype networks, haplotypes of both species (*D. longispina*: four haplotypes, *D. galeata*: seven haplotypes) formed two distinctive clusters, separated by at least 37 mutational steps (Figure C1-2A), in conformity with PCoA results. By contrast, candidate gene haplotype networks all showed a random distribution of haplotypes without any clustering according to species affiliation (*TF*: Figure C1-3; *EA* & *M*: Additional file C1-1: Figures S1 & S2). Species assignment results (Additional file C1-2: Table S2) for *TF* showed that six haplotypes were

found only in *D. galeata*, three only in *D. longispina*, one in both species and their interspecific hybrids and one in *D. galeata* and interspecific hybrids. Haplotype networks thus illustrate that intraspecific divergence is larger than interspecific divergence in several cases.

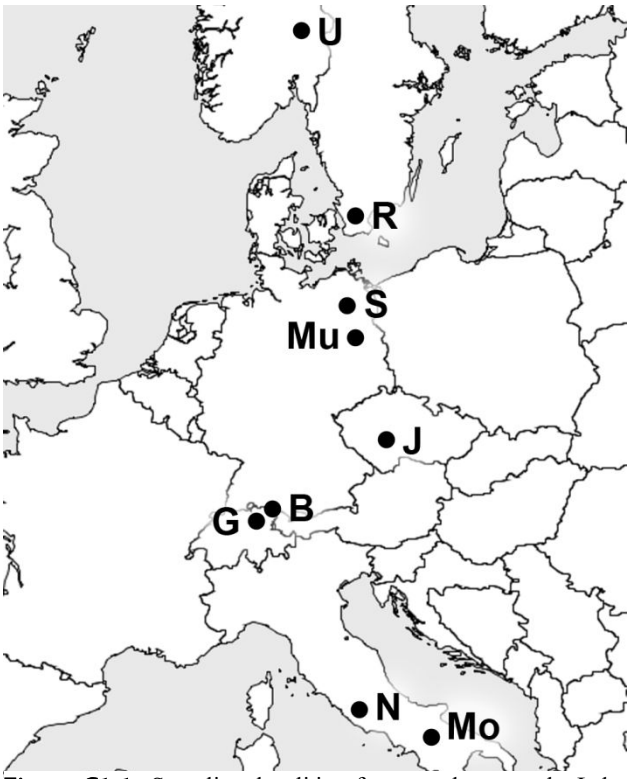

Figure C1-1: Sampling localities from north to south: Lake Ulvevatn (U; Norway), Lake Ringsjön (R; Sweden), Lake Stechlinsee (S; Germany), Lake Müggelsee (Mu; Germany), Jordán reservoir (J; Czech Republic), Lake Constance (B; Austria/Germany/Switzerland), Lake Greifensee (G; Switzerland), Lago di Nemi (N; Italy), and Lago di Monticchio (Mo; Italy)

Since differentiation among populations might have been overruled by a phylogenetic signal of two species, two two-leveled AMOVAs were conducted based only on individuals belonging to one of the parental species. *D. galeata* and *D. longispina* individuals were either grouped according to their taxonomic affiliation (thereafter called "species-AMOVA"), or according to localities (called "locality-AMOVA"). The "species AMOVAs" (Table C1-2) revealed that a major share of variation at loci *TF* and *EA* was observed within the species (77.6 % and 82.5% of total variation respectively) and not between them (22.4% and 17.5% respectively; *TF*: F_{ST} = 0.224; $P < 0.0001$; df = 1; *EA*: F_{ST} = 0.599; $P < 0.0001$; df. = 1), whereas variance was distributed more evenly at locus *M* (41.8% of total variation observed within vs. 58.2% between species, F_{ST} = 0.582; $P < 0.0001$; df = 1). According to the "locality-AMOVAs" (Table C1-2A), based on locality-grouped data, a higher share of variation at loci *TF* and *M* was observed among (62.0% and 56.2% of total variation respectively) than within localities (38.0% and 43.8% respectively; *TF*: F_{ST} = 0.620; $P < 0.0001$; df = 5; *M*: F_{ST} = 0.562; $P < 0.0001$; df = 4). By contrast, locus *EA* exhibits the highest partition of variation within localities (43.2% among vs. 56.8% within localities; F_{ST} = 0.432; $P < 0.0001$; df = 5).

Table C1-1 Sampling locations and overview of molecular markers, population sizes and genetic analyses

Location	Country	Coordinates	Sediment layers used	Temp	Pop code	No. of analyzed individuals					Analyses
						µsats	*EA*	*M*	*TF*	*12S*	
Lake Constance	Germany	47° 37' 21'' N 9° 26' 24'' E	2000-2010 1970-1980		B1 B2	38 45	0 0	0 0	0 0	0 0	1 1
Lake Greifensee	Switzer-land	47° 21'20" N 8° 40'10" E	2000-2010 1973-1983	9.26 °C 8.11 °C	G1 G2	40 36	7 10	15 14	11 15	5 0	1¸ 2, 3, 4, 5, 6, 7 1, 2, 4, 5, 6
Lake Stechlinsee	Germany	53° 9'6" N 13° 34'0" E	1994-2010 1970-1986 1942-1962	9.57 °C 10 °C* 8.53 °C	S1 S2 S3	39 57 50	13 11 9	14 16 15	0 15 14	0 4 2	1¸ 2, 3, 4 1¸ 2, 4, 5 1¸ 2, 4, 5
Lago di Monticchio	Italy	40° 55'57" N 15° 36'17" E	2000-2010	10.14 °C	Mo	27	13	14	18	12	1¸ 2, 3, 5, 6, 7
Lago di Nemi	Italy	41° 42'52" N 12° 42'18" E	2000-2010	15.90 °C	N	44	14	10	15	16	1¸ 2, 3, 5, 6, 7
Jordán reservoir	Czech Republic	49° 24'55" N 14° 39'49" E	2000-2010 1970-1980	12.56 °C	J1 J2	38 36	14 0	13 0	16 0	14 0	1¸ 2, 3, 5, 6, 7 1
Lake Müggelsee	Germany	52° 26'6" N 13° 38'6" E	2000-2010 1975-1985	8.75 °C	Mu1 Mu2	44 31	13 0	0 0	14 0	6 0	1, 2, 3, 6, 7 1
Lake Ringsjön	Sweden	55° 52'1" N 13° 31'25" E	2005-2010		R	45	0	0	0	0	1
Lake Ulvevatn	Norway	59° 60'2"2 N 11° 67'3" E	2005-2011 1975-1985		U1 U2	25 21	0 0	0 0	0 0	0 0	1 1
Total					16	616	104	111	118	59	

Pop code: population code; Temp: average annual air temperature near surface (in populations with candidate gene data).

For analyses based on microsatellite data, neutrality was determined by outlier tests and for analyses based on specific taxa, taxon affiliation was determined by NewHybrids analyses.

* corrected value accounting for artificial warming

1: PCoA and NewHybrids analyses based on all individuals.

2: Haplotype network reconstruction and "species-" as well as "locality-AMOVA" based on *D. galeata* and *D. longispina* individuals successfully genotyped at both the *12S* locus or candidate gene and microsatellite loci; estimation of F_{ST}- and F'_{ST}-values, candidate loci diversity indices, identification of amino acid substitutions between haplotypes, Mantel tests and partial Mantel tests for correlation with temperature as well as tests for species-wise HWE based on *D. galeata* and *D. longispina* microsatellite data respectively.

3: Mantel tests and partial Mantel tests for correlation with geographic distance based on all candidate marker sequences and neutral microsatellite data independent of taxonomic affiliation; "population-AMOVA" based on all candidate marker sequences independent of taxonomic affiliation.

4: "times-AMOVA" based on all candidate marker sequences and microsatellite data independent of taxonomic affiliation

5: LOSITAN and BayeScan outlier tests including all individuals based on all candidate marker sequences and microsatellite data independent of taxonomic affiliation.

6: LOSITAN and BayeScan outlier tests including solely *D. galeata* individuals based on all *D. galeata TF* candidate marker sequences and microsatellite data; Mantel tests and partial Mantel tests for correlation with temperature as well as estimation of F_{ST}- and F'_{ST}-values including solely *D. galeata* individuals based on all *D. galeata TF* candidate marker sequences and neutral microsatellite data.

7: Mantel tests and partial Mantel tests for correlation with geographic distance including solely *D. galeata* individuals based on *D. galeata TF* candidate marker sequences and neutral microsatellite data.

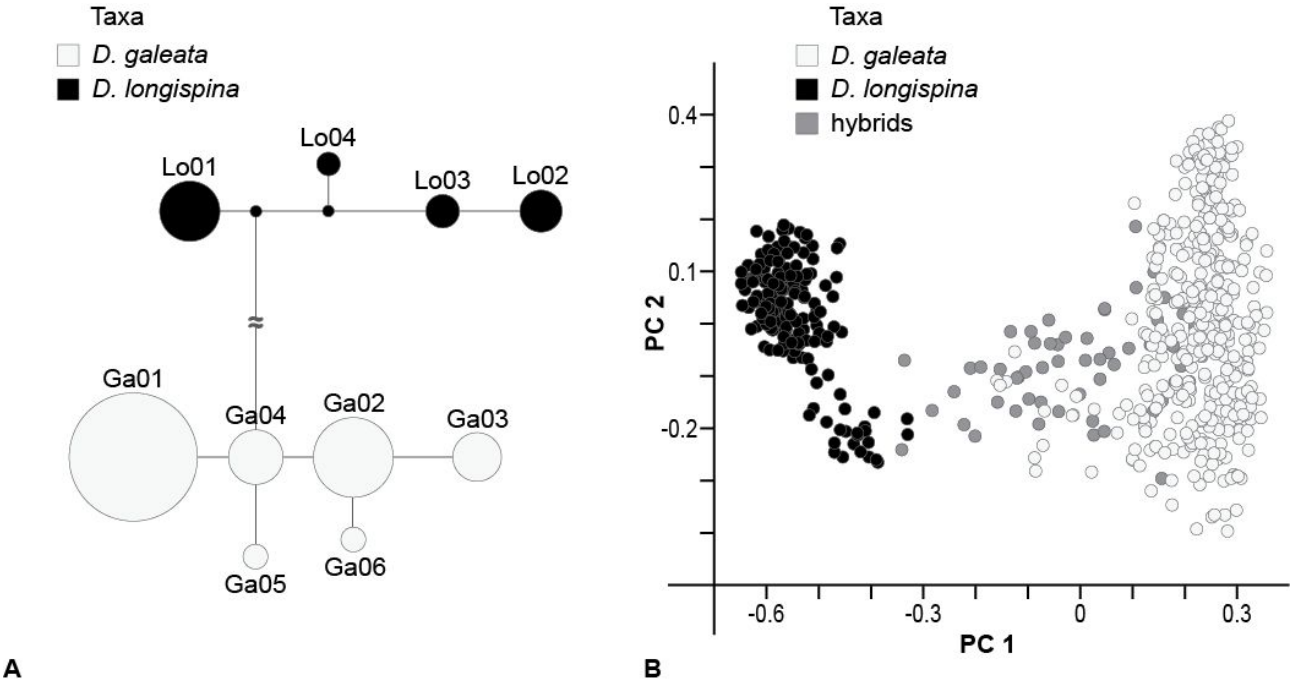

Figure C1-2: Taxon partition at mitochondrial and microsatellite loci. A: Median joining haplotype network, reconstructed using mitochondrial *12S* rDNA sequences. The size of circles is proportional to haplotype frequency. The smallest circle represents one haplotype. The length of lines between circles is proportional to mutational distance; the shortest distance represents one mutational step. The longest distance, indicated by the intercepted line, corresponds to 37 mutational steps. The black dots represent median vectors. B: Principal coordinates analysis of microsatellite data. PC 1 explains 55.0% and PC 2 12.5% of the total variance. Taxon affiliation according to NewHybrids analyses

Genetic differentiation across space, time, and species borders

Microsatellite and candidate locus F_{ST}-values were computed for all individuals as a measure for genetic differentiation among populations (i.e., localities) and subpopulations (i.e., time periods). Additionally, *TF* and microsatellite F_{ST}-values were computed for solely *D. galeata* individuals to assess the bias introduced by analyzing a species complex. This was not possible for loci *EA* and *M* because of limited sequencing success in *D. galeata* clones (from populations J1 and G1). We also computed standardized F_{ST}-values (F'_{ST}, Hedrick 2005) to allow for a comparison of genetic differentiation across different loci.

Although allopatric populations (*D. galeata*: Lago di Nemi and Lago di Monticchio; *D. longispina*: Lake Stechlinsee) exhibit the highest F_{ST}- and F'_{ST}-values in pairwise comparisons at loci *M* and *TF* (e.g., Lago di Nemi versus Stechlinsee 3: F_{ST} (*M*) = 0.681; F_{ST} (*TF*) = 0.822), a general inflation of candidate F_{ST}-values can be rejected at least for locus *TF*. Candidate markers exhibited higher F_{ST}-values than microsatellites (highest value 0.874 at *TF*, Lake Greifensee 1 versus Lago di Nemi) in analyses of all individuals and, importantly, the range of F_{ST}-values at candidate locus *TF* remained stable when analyzing solely *D. galeata*. Microsatellite F_{ST}-values based on solely *D. galeata* individuals by contrast were substantially lower (highest value: 0.124, Lake Müggelsee versus Lago di Nemi) than F_{ST}-values based on all individuals (spanning from zero among the Lake Stechlinsee

subpopulations to 0.392 between S1 and Lago di Nemi; for details on F_{ST}-values see Additional file C1-1: Results and Additional file C1-3: Tables S3-S8).

To test if genetic differentiation in general is higher over time or rather over space, another, three-leveled AMOVA (thereafter "times-AMOVA") was conducted. The "times-AMOVA" included all individuals (independent of taxonomic affiliation) from localities with different time periods (Table C1-1) and revealed a much lower genetic differentiation over time than over space (all loci: F_{CT}-values < 0; df = 1; Table C1-3).

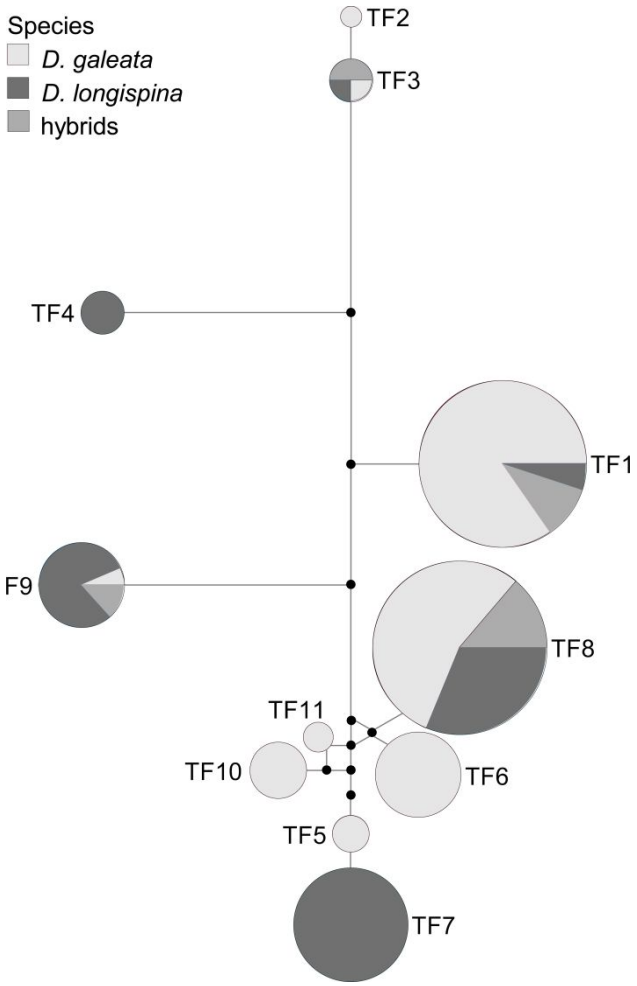

Figure C1-3: Median joining haplotype networks reconstructed using *TF* candidate gene marker sequences. The size of circles is proportional to haplotype frequency. The smallest circle represents one haplotype. The length of lines between circles is proportional to mutational distance; the shortest distance represents one mutational step. The black dots represent median vectors. Indels were considered a single mutational event. A: The coloring corresponds to the taxa (taxon affiliation according to NewHybrids analyses). B: The coloring corresponds to the populations

Table C1-2 Two-leveled AMOVAs. A: species- and locality-based AMOVAs; B:) "population-AMOVAs"

Locus	Sites included	Level	df	F_{ST}	Variation among [%]	Variation within [%]
A						
EA	All	Species[1]	1	0.172*	17.2	82.8
	All	Locality[1]	5	0.288*	28.8	71.2
M	All	Species[1]	1	0.582*	58.2	41.8
	All	Locality[1]	4	0.562*	56.2	43.8
TF	All	Species[1]	1	0.224*	22.4	77.6
	All	Locality[1]	5	0.620*	62.0	38.0
B						
EA	All	Population[2]	5	0.33369*	33.4	66.6
	Exons	Population[2]	5	0.29671*	29.7	70.3
	Introns	Population[2]	5	0.36128*	36.1	63.9
M	All	Population[2]	4	0.33962*	34	66
	Exons	Population[2]	5	0.23408*	23.4	76.6
	Introns	Population[2]	5	0.34271*	34.3	65.7
TF	All	Population[2]	4	0.52392*	52.4	47.6
	Exons	Population[2]	4	0.36356*	36.4	63.6
	Introns	Population[2]	4	0.56445*	56.4	43.6

Species: Individuals assigned one of two groups according to their taxonomic affiliation, Locality: Individuals assigned to groups according to their lakes of origin (G1 and G2 form one single group as well as S1, S2, and S3), Population: Individuals assigned to groups according to their population.

[1] Only individuals assigned to one of the parental species (based on NewHybrids analyses) were included in these analyses.

[2] Only individuals from 2000-2010 (independent of species affiliation) were included into the analyses.

* $P < 0.0001$

Table C1-3: Three leveled "times-AMOVA"

	among times			among populations within times			within populations	among populations
Locus	df	F_{CT}	Var. [%]	df	F_{SC}	Var. [%]	Var. [%]	F_{ST}
EA	1	-0.128	-12.78	2	0.265*	29.87	82.91	0.171
M	1	-0.235	-23.49	2	0.384*	47.38	76.10	0.239
TF	1	-0.125	-12.49	2	0.250*	28.11	84.38	0.156
sats	1	-0.200	-19.98	2	0.210*	40.95	79.03	0.341

Variance between groups of populations, between populations within groups and within populations. Only individuals from localities with samples from different times were included into the analysis and grouped according to the time period they are from.

sats: microsatellite loci; df: degrees of freedom; Var.: variation

*: $P < 0.0001$

Neutral versus functional genetic differentiation

For a direct comparison of genetic differentiation among populations between functional (ecologically relevant) and neutral markers in spite of different levels of heterozygosity, F'_{ST}-values were compared. In a comparison of F'_{ST}-values including all individuals and loci, we observed that all three candidate loci deviated from neutral expectations as defined by microsatellite F'_{ST}-values (Additional file C1-1: Figure S3 and Additional file C1-3: Tables S9-S12). Notably, standardized F'_{ST}-values between populations at loci *EA* and *M* were lower than the values inferred from microsatellites (i.e., neutral loci) in almost every comparison By contrast, at locus *TF*, standardized genetic differentiation between populations was higher than the differentiation at neutral loci in 55% of the comparisons; thus indicating directional selection and local adaptation. Differentiation over time in Lake Greifensee was nearly equally high at neutral microsatellite loci and candidate locus *TF* (F'_{ST} (*TF*) = 0.162; F'_{ST} (microsatellites) = 0.158).

Including a species complex in an analysis of locus neutrality may bias the results, or phylogenetic signals may obscure patterns of adaptation. To account for that, the comparison of neutral and functional genetic differentiation was repeated for locus *TF* including solely *D. galeata* individuals (Additional file C1-1: Figure S4 and Additional file C1-3: Tables S13 and S14). This resulted in a notably increased number of pairwise comparisons between populations that diverged from neutral expectations compared to the analysis including both species and interspecific hybrids. Again, there was no indication for an inflation of candidate gene F'_{ST}-values but instead for microsatellite F'_{ST}-values: Standardized differentiation over time at locus *TF* in Lake Greifensee was constant regardless whether only *D. galeata* or all individuals where included, whereas at microsatellite loci, it was much lower for solely *D. galeata* (F'_{ST} (*TF*) = 0.166; F'_{ST} (microsatellites) = 0.092).

Outlier tests for natural selection

To statistically test for locus neutrality, a coalescent-based simulation method (LOSITAN, Antao *et al.* 2008) as well as a Bayesian approach (BayeScan, Foll & Gaggiotti 2008) were used. Both tests were conducted on two data sets consisting of 11 microsatellite loci and three candidate gene markers each and containing either all individuals or only *D. galeata*. For both outlier tests, it was necessary to reduce the candidate marker sequence data to genotypic data. This reduction inevitably biased the genetic distances between the haplotypes. To correct for this bias, we used the method of Eveno *et al.* (2008) and pooled haplotypes with less than 10% sequence divergence (arbitrary threshold, for details see Additional file C1-3: Table S15). This resulted in a maximum number of 16, 39, and 7 mutations between pooled sequences in *EA*, *M*, and *TF*.

Outlier tests revealed no significant signal of natural selection for any of the candidate loci when the unpooled data set including all individuals was analyzed (Figure C1-4A and C). In the LOSITAN analysis, microsatellite loci *Dp281NB* and *Dgm109* were detected as outliers exceeding the 95% and 1% confidence interval for neutral differentiation respectively. Analyses of the pooled data set containing all individuals both identified *TF* as an outlier (exceeding the 99% confidence

interval in LOSITAN and the 95% critical value for neutral differentiation in BayeScan); thus suggesting directional selection acting on this candidate locus (Figure C1-4B and D). *EA* and *M* showed no significant deviation from neutral expectations. Putatively neutral microsatellite loci *Dp281NB* (exceeding the 95% confidence interval), *SwiD6*, *SwiD18* (exceeding the 5% confidence interval), and *Dgm109* (exceeding the 1% confidence interval) were identified as outliers by LOSITAN but not by BayeScan.

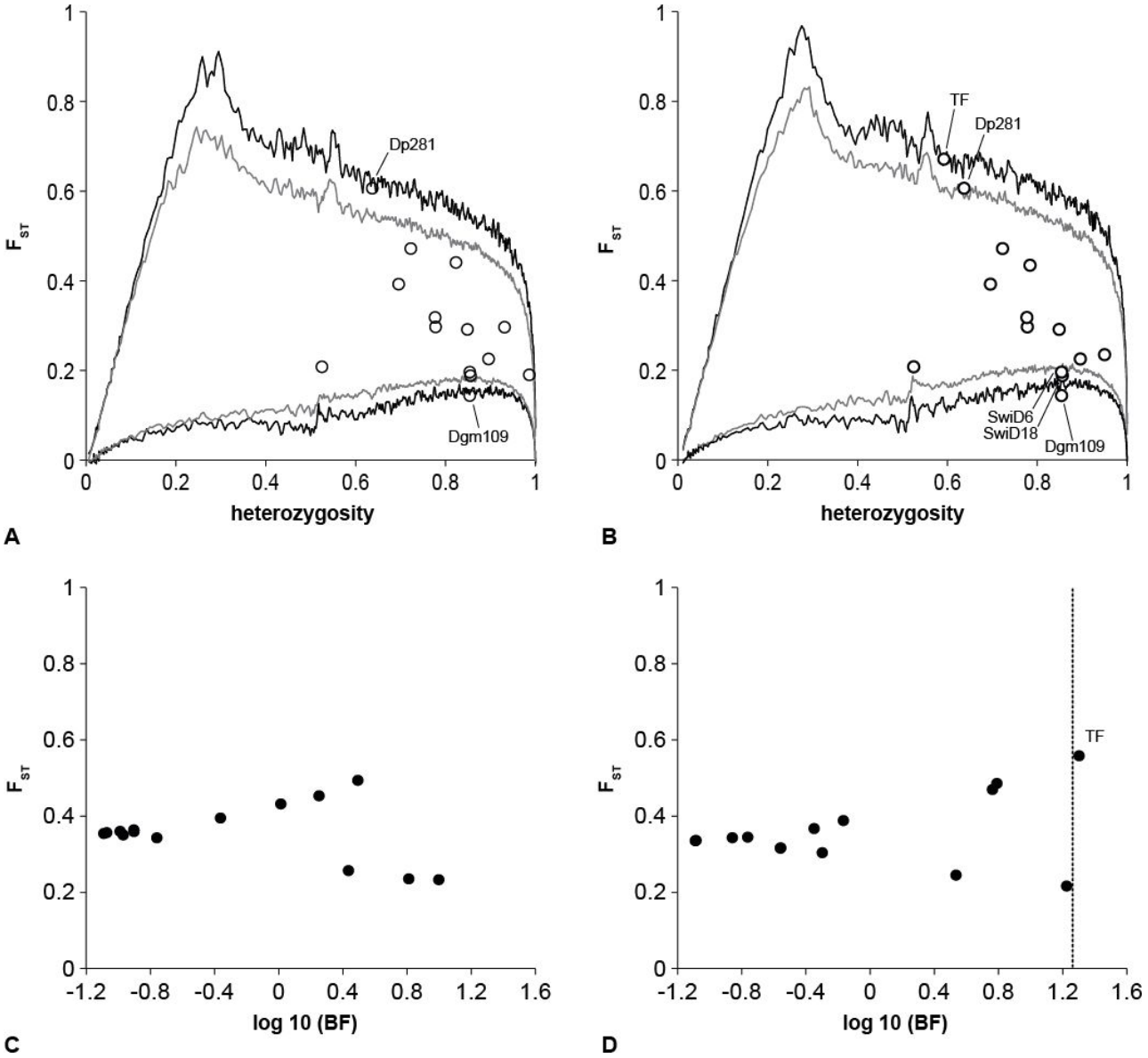

Figure C1-4: Outlier tests including all taxa. A and B: LOSITAN. The dotted lines indicate the 95% and 5% confidence limits, the solid lines the 99% and 1% confidence limits of the simulated distribution of F_{ST}-values under neutrality. C and D: BayesScan. The F_{ST} coefficients averaged over populations are plotted against their probability (*P*-values) for being outliers. The dashed line represents the 95% critical value for directional selection. PO = 4. A & C: based on unpooled data, B & D: haplotypes with < 10% sequence divergence were pooled

Analyzing solely *D. galeata* individuals in LOSITAN (Additional file C1-1: Figure S5) yielded the same result using both the pooled and the unpooled data set: candidate locus *TF* and microsatellite locus *SwiD14* were identified as outliers under directional selection (outside the 99% and 95% confidence interval respectively). None of the loci was identified as an outlier by BayeScan.

Isolation by distance and environment

Mantel tests were used to infer correlations between (standardized) genetic differentiation (F_{ST}- and F'_{ST}-values) and geographic distance (a test for isolation by distance, IBD, in case of neutral markers) and between genetic differentiation and temperature differences. To control for temperature differences while testing for a correlation with geographic distance and vice versa, partial Mantel tests were used.

Standardized genetic differentiation and geographical distance were significantly correlated at locus *M* (F_{ST}: Mantel test $R = 0.859$, $P > 0.05$; F'_{ST}: Mantel test $R = 0.966$, $P < 0.05$; Table C1-4). After correcting for temperature differences, the correlation with geographic distance was significant for both standardized and non-standardized genetic differentiation (F_{ST}: partial Mantel test $R = 0.868$; $P < 0.05$, F'_{ST}: partial Mantel test $R = 0.966$; $P < 0.05$; Table C1-4). *EA* and *TF*, and microsatellite loci showed no significant correlation of genetic differentiation and geographic distance; therefore no indication for IBD among populations was found.

All analyses revealed significant correlations between pairwise differences in average annual air temperatures among sampling sites and pairwise genetic differentiation among populations at candidate locus *TF*. Correcting for the geographical distance resulted in increased, highly significant correlation coefficients (F_{ST}: Mantel test $R = 0.641$, $P < 0.01$; partial Mantel test $R = 0.658$; $P < 0.01$; F'_{ST}: Mantel test $R = 0.633$, $P < 0.05$; partial Mantel test $R = 0.642$; $P < 0.01$; see Table C1-4). None of the other markers showed a significant correlation between temperature and genetic differentiation.

Table C1-4: Mantel and partial Mantel test results

	Correlation coefficients (*R*) F_{ST}				Correlation coefficients (*R*) F'_{ST}			
	EA	*M*	*TF*	sats	*EA*	*M*	*TF*	sats
temp	0.188	0.176	0.641**	0.212	0.237	0.076	0.633*	0.051
geo	-0.259	0.859	-0.063	0.275	-0.248	0.966*	0.279	0.108
temp x geo	0.252	0.164	0.190	0.252	0.252	0.164	0.190	0.252
PC temp (geo)	0.173	0.064	0.658**	0.096	0.216	-0.278	0.642**	-0.074
PC geo (temp)	-0.279	0.868*	-0.055	0.263	-0.273	0.966*	0.336	0.106

Temp: temperature difference; geo: geographic distance; PC: partial correlation (constant parameter in brackets). Only neutral microsatellite loci (determined by outlier tests) were analyzed.

*: $P < 0.05$; **: $P < 0.01$

Mantel tests and partial Mantel tests for candidate locus *TF* and selectively neutral microsatellite loci were additionally conducted based on solely *D. galeata* values to avoid a bias potentially introduced by analyzing a species complex. Mantel tests showed a significant correlation only for F'_{ST} and temperature differences at locus *TF* (F_{ST}: Mantel test $R = 0.583$, $P > 0.05$; Additional file C1-3: Table S16), whereas partial Mantel tests revealed significant correlations between both F_{ST} and F'_{ST}, and temperature differences (partial Mantel test $R = 0.585$, $P < 0.05$; F'_{ST}: Mantel test $R = 0.567$, $P < 0.05$; partial Mantel test $R = 0.565$, $P < 0.05$). No correlations were found at microsatellite loci and with geographic distance.

Comparisons of coding and non-coding regions

To test whether the amount of variation among populations is higher than within (as expected if local adaptation occurred), two-leveled "population-AMOVAs" were conducted. Therefore, data sets including all individuals (independent of taxonomic affiliation) from the last decade were grouped according to their population affiliation. At loci *M* and *EA*, the majority of the total variance was found within populations (approximately two thirds to one third both; *M*: $F_{ST} = 0.334$; df = 4; $P < 0.0001$; *EA*: $F_{ST} = 0.340$; df = 5; $P < 0.0001$; "population-AMOVA", Table C1-2B). By contrast, at locus *TF* the variance was almost equally distributed among and within populations ($F_{ST} = 0.524$; $P < 0.0001$; df = 4).

Additionally to the complete sequences, exonic and intronic regions were analyzed separately to conclude, whether differentiation among populations is based on divergence of gene products or rather regulatory differences. Similar to the complete sequences, most of intronic as well as exonic variation was found within populations at *EA* and *M* (again approximately two thirds respectively; "population-AMOVA", Table C1-2B). At locus *TF*, the major share of exonic variation was also found within populations (also approximately two thirds), whereas the share of intronic variation was higher among populations (56.4% of total variation; $F_{ST} = 0.564$; $P < 0.0001$; df = 4). The impact of exonic variation was assessed by computing the number of synonymous and nonsynonymous substitutions at each locus, and the ratio ( ) of nonsynonymous substitutions per nonsynonymous site (K_a) to silent substitutions per silent site (K_s). Haplotypes at locus *TF* showed 16 amino acid substitutions, which is the highest number among all loci (*EA*: 11, *M*: five, Additional file C1-3: Table S17). *TF* also showed with = 0.636 by far the highest mean K_a/K_s ratio ($_{EA}$ = 0.091, $_{M}$ = 0.053, Additional file C1-3: Table S18).

Discussion

Patterns of genetic divergence among *D. galeata* and *D. longispina* show a pronounced contrast between neutral and ecologically relevant loci. At neutral mitochondrial and nuclear loci, a distinct differentiation of parental species and an intermediate status of interspecific hybrids was detected, which is in accordance with studies on the phylogeny of the *D. longispina* species complex (Schwenk *et al.* 2000; Petrusek *et al.* 2012) and on population structure in European *Daphnia*

(Thielsch *et al.* 2009). By contrast, candidate loci exhibit a pattern of genetic differentiation that does not discriminate between taxa or reflect strong barriers to gene flow. Clusters in candidate loci-based haplotype networks contain haplotypes of both parental species and shared haplotypes found in both parental species were observed for each candidate locus. In several cases, intraspecific genetic distance among candidate loci haplotypes is larger than interspecific genetic distance. Thus, haplotype clustering is clearly not in concordance with species affiliation, an observation expected under differential introgression among species, but also under incomplete lineage sorting or convergence among taxa (Hedrick 2013).

Considering the abundance of hybridization events among the species of the *D. longispina* species complex (Schwenk 1993; Gießler *et al.* 1999; Taylor *et al.* 2005; Ishida & Taylor 2007b; Dlouha *et al.* 2010; Ishida *et al.* 2011; Thielsch *et al.* 2012), introgression represents the most plausible explanation for the observed patterns. This notion is, at least at loci *M* and *TF*, supported by the large differences in allele frequencies among the allopatric populations of the two species, Lake Stechlinsee (*D. longispina*) and Lago di Nemi/Lago di Monticchio (*D. galeata*) at candidate loci *TF* and *M*. By contrast, at loci exhibiting ancestral polymorphisms and incomplete lineage sorting, we would expect little differentiation among parental species (Larson *et al.* 2013).

Two different outlier tests and the comparative analyses of F'_{ST}-values are in support of selection acting on *TF*. The results of these different methods were compared, as suggested by (Shimada *et al.* 2011), to reduce the occurrence of false positives. In contrast to the microsatellite loci identified as outliers solely by LOSITAN (*Dp281NB*, *SwiD6*, *SwiD18*, and *Dgm109*), *TF* is the only locus showing a consistent pattern over all tests and hence less likely to be a false positive. Candidate loci *EA* and *M* however show very weak signs of selection at most, indicated only by the comparative analyses of F'_{ST}-values. Reducing the data set for locus *TF* dispelled concerns regarding the combined analysis of two parental species and their interspecific hybrids. Despite the reduced sample size and thus decreased statistical power, *TF* was also an outlier when only *D. galeata* individuals were analyzed. Furthermore, the effect of analyzing only *D. galeata* on the mean F'_{ST}-values was negligible in spite of the high amounts of genetic differentiation found between allopatric populations of the parental species. Positive results from outlier tests may arise through other processes besides natural selection. Population structure (Excoffier *et al.* 2009), demographic processes (Bamshad & Wooding 2003; Hofer *et al.* 2009) and background selection against deleterious mutations (Charlesworth *et al.* 1997)) are frequently discussed in the context of genome scan studies, which typically use data mining approaches and thus are prone to false positive results. All of these explanations can be ruled out, as they are expected to have affected the whole genome equally instead of being restricted to specific loci (Bamshad & Wooding 2003). Thus, they are not sufficient to explain the incongruence of markers in PCoA and network analyses.

Candidate genes were chosen based on their experimentally demonstrated response to temperature variation and their role in resistance to cyanobacterial toxins (Agrawal *et al.* 2005; Schwerin *et al.*

2009; Asselman *et al.* 2012; De Coninck *et al.* 2014; Asselman *et al.* 2015). Over time, we found increased temperatures at sampling sites and observed differentiation at candidate loci among Lake Greifensee subpopulations and a discrepancy between microsatellite and *TF* F'_{ST}-values over time. This might reflect direct or indirect selection through increased temperatures. These observations are in accordance with recent literature. Phenotypic adaptive changes as a response to temperature increase have already been found in experimental studies using *D. magna*, including changes in natural populations over a time span comparable to the time span covered in our study (Van Doorslaer *et al.* 2009a; Van Doorslaer *et al.* 2010; Geerts *et al.* 2015). Furthermore, studies on *D. galeata* populations from different locations revealed a divergent phenotypic response to temperature and evidence for temperature-driven micro-evolution over time in Lake Constance (Henning-Lucass *et al.* 2016, N. Henning-Lucass, Personal Communication), matching the pattern observed at candidate loci. No general signal of genetic differentiation among *Daphnia* populations over time however was detected at any locus by the "times-AMOVA". Therefore, we cannot provide evidence for local adaptation at locus *TF* over the observed timespan, possibly because the selective pressure was too weak to promote a divergence pronounced enough to be visible in the AMOVA results.

The association of genetic differentiation with temperature differences revealed by Mantel tests and partial Mantel tests suggests that differing temperatures however might indeed directly or indirectly shape genetic differentiation among populations from different localities at locus *TF*. The association of genetic differentiation and geographical distances found for candidate locus *M* might, by contrast, reflect a weak response to any selective pressure associated with latitude, since the neutral microsatellite loci exhibit no pattern of IBD (which is in agreement with other studies on European *Daphnia* species, e.g., Thielsch *et al.* 2009). The physiological mechanism underlying selection acting on *TF*, the marker within the trypsin-encoding gene *TRY5F*, is possibly linked to varying rates of metabolic activity as a direct consequence of divergent environmental temperatures. In *Daphnia*, accommodating the increased energy demand under elevated temperatures (Paul *et al.* 2004a) is aggravated by the increased frequency of algal blooms (Paerl & Huisman 2008) as a secondary effect of increasing temperatures. Elevated levels of cyanobacteria are problematic for *Daphnia* as they represent a food source of poor quality and produce (among other toxins) serine protease inhibitors. These interfere with trypsins, the most important gene family in the gut of *Daphnia* because of their role in protein digestion and energy supply (together with chymotrypsins, von Elert *et al.* 2004; Agrawal *et al.* 2005). An increased tolerance to cyanobacterial protease inhibitors has been shown on the organismic level as well as on the protein level as a result of the coexistence of *Daphnia* spp. with cyanobacteria (e.g., Hairston *et al.* 1999; Sarnelle & Wilson 2005; Blom *et al.* 2006). Increased tolerance on the protein level was shown to be derived by physiological plasticity, i.e., the expression of different isoforms of digestive enzymes (von Elert *et al.* 2012). In recent studies (Asselman *et al.* 2012; De Coninck *et al.* 2014; Asselman *et al.* 2015), temperature and exposure to different strains or species of cyanobacteria have been

shown to induce differential expression of members of the trypsin gene-family. For locus *TF*, differential selection acting on both the amino acid sequence and on the regulatory level is possible. We inferred non-synonymous substitutions between *TF* haplotypes but on the other hand, the major part of variation among populations originates from the non-coding intronic regions, as revealed by the "population-AMOVAs". An interplay of trypsin expression levels and genetic variation has been previously demonstrated in a study by Schwarzenberger *et al.* (2012), who showed an association of the phylogenetic distance among trypsin alleles, divergent gene expression changes and variation in tolerance to dietary trypsin inhibitors in *D. magna*. However, a variety of other stressors, including salinity (Latta *et al.* 2012), a flame retardant (HCCPD, Houde *et al.* 2013), cadmium (De Coninck *et al.* 2014), and changes in phosphorous content of the diet (Jeyasingh *et al.* 2011) also induced differential expression of trypsin genes. Hence, differential expression of trypsins in general is commonly involved in stress responses. At the same time, the set of differentially expressed trypsin genes varies among stressors as well as among *Daphnia* clones. Differences have also been found in the direction of expression. These observations demonstrate specificity in the expression of trypsins in response to different stressors (contrasting with an unspecific stress response that would be expected not to vary among different stressors).

We propose that local selection-driven differential introgression most likely explains the elevated levels of genetic differentiation observed among populations at locus *TF* and possibly promoted the deviance from patterns found at presumably neutral, non-functional microsatellite loci. This conclusion is supported by the distribution of variance among the taxa and the sampling localities (evaluated in "species-" and "locality-AMOVAs"), which clearly showed that a major share of variation at locus *TF* is explained by the locality rather than the taxonomic background and substantial differentiation at *TF* is found among localities, not among species. Comparative analyses of F'_{ST}-values among Lake Greifensee subpopulations revealed that although we found genetic differentiation over time at microsatellite markers in Lake Greifensee, the level of genetic differentiation was low. This possibly indicates genetic influx from invading genotypes, but certainly no displacement of the local population. Thus, the levels of gene flow from other populations into Lake Greifensee was low or locally adapted lineages outcompeted invaders (De Meester *et al.* 2002). Analyzing solely *D. galeata* individuals elevated the divergence of F'_{ST}-values between markers among times in Lake Greifensee even further by substantially decreasing microsatellite F'_{ST}-values. This inflation of neutral but not of functional population differentiation by analyzing two species and their interspecific hybrids fits well within the scenario of differential introgression. Genome-scale studies comparing patterns of divergence within hybrid species complexes also found positive correlations between high locus-specific differentiation among parental species and introgression in manakins (Parchman *et al.* 2013) and butterflies (Gompert *et al.* 2012). However, a study on chickadees (Taylor *et al.* 2014) revealed low introgression for highly divergent loci and in all three studies, these associations are only partial. Likewise, examples for highly divergent loci among parental species with both neutral and elevated introgression levels

were shown in spruce (Hamilton *et al.* 2013a, b) and crickets (Larson *et al.* 2013; Larson *et al.* 2014). These inconsistent results indicate the complexity and specificity of these patterns throughout the genome rather than supporting or contradicting our conclusions. A more direct comparison with our results is difficult, because in all studies cited above, locus-specific genetic divergence was estimated solely among species.

To provide further evidence for the potentially important role of candidate gene *TRY5F*, it is necessary to elucidate if and how selection at this locus contributes to local adaptation. Therefore, it is necessary to establish genotype-phenotype relationships at locus *TF* in response to different temperatures and cyanobacteria strains. Furthermore, the genome-wide extent and direction of introgression need to be quantified and introgression events need to be dated using ancient *Daphnia* propagules.

Conclusions

Our study demonstrates the utility of combining analyses of ecologically relevant candidate genes and selectively neutral markers to reconstruct patterns of natural selection (and potentially adaptation) in natural populations. We report an unexpected and striking discrepancy among markers between taxa of the *D. longispina* species complex. Together with the evidence for selection-driven population differentiation at candidate locus *TF*, this suggests the occurrence of introgression at loci under natural selection, a prerequisite for facilitated adaptation by capture of preadapted alleles (although we cannot completely reject of the alternative explanations of convergence and incomplete lineage sorting). While an adaptational advantage though differential introgression might contribute to the rapid local adaptation processes and the persistent permeability of reproduction barriers in *Daphnia* species complexes, further studies are needed to experimentally support this hypothesis.

Methods

Sampling and DNA preparation

Sediment cores were collected from nine freshwater lakes distributed across Europe along a latitudinal gradient from southern Italy to Norway: Lago di Monticchio (Mo, Italy), Lago di Nemi (N, Italy), Lake Greifensee (G, Switzerland), Lake Constance (B, Austria/Germany/Switzerland), Lake Stechlinsee (S, Germany), Lake Müggelsee (Mu, Germany), Jordán reservoir (J, Czech Republic), Lake Ringsjön (R, Sweden), and Lake Ulvevatn (U, Norway); see Figure C1-1. Candidate markers were sequenced for six lakes (Lago di Monticchio, Lago di Nemi, Lake Greifensee, Lake Stechlinsee, Lake Müggelsee, and Jordán reservoir), whereas populations from all nine lakes were genotyped at microsatellite loci. Microsatellite and mtDNA data from sites lacking candidate marker information were solely used for PCoA and NewHybrids analyses (described below) in order to obtain detailed information about population structure and taxon affiliation.

The sediment cores were collected between 2010 and 2011 in deep water sites, where the ephippia density is highest (Caceres & Hairston 1998) and the turbulence lowest. The sediment was divided into 0.5 cm to 1 cm slices and stored at 4 °C in opaque sealed tubs to avoid hatching stimuli. Sediment cores were dated using the methods described in Brede *et al.* (2009). For all lakes, sediment layers corresponding to the last decade (2000-2010), and when possible also from around 1970 to 1980, were pooled (Table C1-1). Lake Stechlinsee experienced a period of artificial warming from 1966 to 1990, when it received warm water discharge from a power plant (described in Casper & Koschel 1995). To cover the periods before, during and after this period, we also used sediment from the years 1942 to 1962.

We used air temperatures as a proxy for water temperatures, as lake surface temperatures from large and deep lakes correspond closely to air temperature (Livingstone & Lotter 1998). The average annual air temperatures near the sampling sites were obtained from the Climate Research Unit of the British Atmospheric Data Centre (Mitchell & Jones 2005; BADC 2008). During the period of artificial warming in Lake Stechlinsee, water surface temperatures increased by 1.1 °C to 10 °C, depending on the distance to the point of warm water discharge. That was taken into account by increasing the temperature values for that period by 1.5 °C, which is a rough estimate of the actual temperature increase (Casper & Koschel 1995).

Ephippia were collected using a 125 µm sieve (Retsch, Haan, Germany) and stored in 70-100% ethanol until the resting eggs were isolated. Resting eggs where extracted from the ephippial case and the DNA prepared from single resting eggs according to Montero-Pau *et al.* (2008). DNA was stored at 4 °C or -20 °C.

Selection of candidate genes

Genes differentially expressed as shown in a proteome study by Schwerin *et al.* (2009) were chosen as candidate genes. Amplification and sequencing was attempted for DNA sequences of all eleven differentially expressed proteins therein and successful for the three loci *EA*, *M*, and *TF* within the genes *ERNA*, *MHC-1*, and *TRY5F*. Primer pairs were designed using the available DNA sequences of *D. pulex* (Colbourne *et al.* 2011). Details on all potential candidate genes are listed in Additional file C1-3: Table S19.

Microsatellite analyses and DNA sequencing

Eleven microsatellite loci (*DaB10/14*, *DaB17/17*, *Dp519*, *Dgm105*, *Dgm109*, *Dgm112*, *Dp281NB*, *SwiD14*, *SwiD6*, *SwiD12,* and *SwiD18*; Brede *et al.* 2006) were analyzed for each individual according to the protocols in Thielsch *et al.* (2012) for multiplex PCR, fluorescence labeling, and DNA sequencing. Microsatellite data were verified using MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.* 2004; applying Bonferroni correction and 1000 bootstrap replicates) and tested for Hardy-Weinberg equilibrium (HWE) in Arlequin ver. 3.5.1.2 (Excoffier & Lischer 2010). None of the microsatellites showed evidence for scoring errors, large allele dropout, and null alleles

according to rank-based probabilities of observed homozygote frequencies. The species-wise tests showed that one locus deviated from HWE (*Dgm109* in 40% of populations) in *D. longispina* (mean = 96.4%, sd = 11.5%). In *D. galeata*, no locus deviated from HWE in more than 69.2% (*Dgm109*) of populations (mean = 90.4%, sd = 11.3%). Bonferroni correction was applied and monomorphic loci were excluded from HWE tests.

Mitochondrial *12S* rDNA was sequenced for 59 individuals and the three nuclear loci *M*, *EA*, and *TF* for five to 15 randomly selected individuals per population (Table C1-1). For details on PCRs and primer sequences see Additional file C1-3: Table S20. PCR products were purified using the Agencourt AMPure Kit (Beckman Coulter, Brea, CA, USA). To validate the sequence data, loci *EA* and *M* were cloned from 30 and 12 individuals respectively (detailed description in Additional file C1-1: Methods). Sequencing was performed on an ABI 3730 capillary sequencer (Life Technologies, Carlsbad, CA, USA).

Sequences were aligned and manually adjusted using Geneious v.5.0.3 (Drummond *et al.* 2011) and forward and reverse sequences when possible. Superposition of bases was regarded as heterozygous when found in both the forward and the reverse sequence. Sequences were reconstructed from superimposed chromatograms when one of the sequences corresponded to a homozygous, cloned or previously reconstructed sequence and sequences with superimposed chromatograms of more than two sequences were discarded. Introns and exons were assigned by aligning the sequences to the *D. galeata* transcriptome, generated through 454 Roche sequencing (Daphnia Genomics Consortium, data available upon request).

Data analysis

Species affiliation and haplotype distribution

Microsatellite genotype data were analyzed for species affiliation and identification of interspecific hybrids according to Thielsch *et al.* (2012). We used the program NewHybrids v1.1 (Anderson & Thompson 2002) which utilizes a Bayesian approach to assign individuals to one of two parental species, the F1 generation or a backcross generation. We performed a run on a data set containing microsatellite data from all 16 populations and an additional *D. galeata* reference data set from the study of Thielsch *et al.* (2009) using 50000 burn-in sweeps, 100000 sweeps after burn-in and Jeffreys prior. A covariance-standardized principal coordinate analysis (PCoA) based on the same data set was then conducted in GenAlEx 6 (Peakall & Smouse 2006; Additional file C1-2: Table S1).

To analyze intra- and interspecific relationships among haplotypes, we generated haplotype networks using the Median-Joining method (Bandelt *et al.* 1999) implemented in Network 4.611 (www. fluxus-engineering.com) for *12S* rDNA as well as candidate loci sequences. Only individuals genotyped both at candidate gene and microsatellite loci were used to generate haplotype networks and indels were considered as a single mutational event.

Analyses of molecular variance

Arlequin ver. 3.5.1.2 was used for estimation of diversity indices, analyses of molecular variance (AMOVAs) and calculation of F_{ST}-values for candidate loci. The haplotype level option was used for candidate markers, the locus level option for microsatellites and default settings for all loci. Genetic distances between haplotypes were computed based on AMOVAs (carried out using 10000 permutations and the locus-by-locus option) and the Kimura 2 parameter method (for candidate loci) or the number of different alleles (for microsatellites). Microsatellite codom-allelic F_{ST}-values were computed in GenAlEx. Bonferroni correction was used to correct all F_{ST} *P*-values for multiple testing. Microsatellite and *TF* F_{ST}-values were computed for all and for solely *D. galeata* individuals. *EA* and *M* F_{ST}-values were only computed for all individuals because of the limited sequencing success in *D. galeata* (of populations J1 and G1). Only selectively neutral microsatellite loci (determined by outlier tests) were used to compute F_{ST}-values, reducing the number of microsatellite loci to seven (*DaB10/14*, *Dgm105*, *Dgm112*, *Dp519*, *DaB17/17*, *SwiD12*, and *SwiD14*, all individuals) and ten or six (*D. galeata* only). *D. galeata* microsatellite F_{ST}-values were computed twice. First, *SwiD14*, the non-neutral locus inferred from *D. galeata*-based outlier tests was excluded. Second, loci inferred to be non-neutral in outlier tests on all individuals were excluded as well (*SwiD14*, *SwiD6*, *SwiD18*, *Dg281*, and *Dgm109*). F_{ST}-values differed only marginally between these two analyses; thus only values based on ten microsatellite loci are shown (F_{ST}-values based on six microsatellite loci are available upon request).

Four different AMOVA analyses were carried out using different subsets of the data. Since differentiation among populations might have been obscured by a phylogenetic signal, two two-leveled AMOVAs were conducted based on individuals belonging to one of the parental species (identified via the NewHybrids analyses) and excluding interspecific hybrids and individuals lacking microsatellite data. *D. galeata* and *D. longispina* individuals were either grouped according to their taxonomic affiliation ("species-AMOVA"), or according to localities ("locality-AMOVA"). Thereby individuals from different times were grouped according to localities. A third two-leveled AMOVA ("population-AMOVA") was conducted to test whether the amount of variation among populations is higher than within (as expected if local adaptation occurred), regardless of their taxonomic affiliation. All individuals (independent of taxonomic affiliation) from 2000-2010 were used and grouped according to their population affiliations. "Population-AMOVAs" were carried out using the complete sequences as well as only intronic and exonic sites respectively. To show if genetic differentiation is higher over time or rather over space, a fourth, three-leveled AMOVA ("times-AMOVA") was conducted. All individuals (independent of taxonomic affiliation) from localities with different time periods were included in the analysis and grouped accordingly (*EA*, *M*, microsatellites: group 1: G1 & S1, group2: G2 & S2; *TF*: group 1: G1 & S2, group 2: G2 & S3; Table C1-1).

The number of synonymous and nonsynonymous substitutions at each locus, the number of silent substitutions per silent site (K_s) and nonsynonymous substitutions per nonsynonymous site (K_a) were computed using DnaSP v5 (Librado & Rozas 2009).

Tests for marker neutrality based on genetic differentiation

The program GenoDive (Meirmans & Van Tienderen 2004) was used to assess standardized population differentiation (F'_{ST}; Hedrick 2005) at candidate gene and microsatellite markers in two data sets, either for all individuals or exclusively for *D. galeata* individuals precisely as described for F_{ST}-values above. The standardized values of population differentiation enabled us to compare population structuring at putatively neutral microsatellite and functional candidate gene markers in spite of different levels of heterozygosity (Hedrick 2005). *D. galeata* microsatellite F'_{ST}-values based on ten and six microsatellite loci were also compared (see definition of the loci sets in the previous section). Results of both data sets differed only marginally and showed deflation of F'_{ST}-values compared to analyses including all individuals. Therefore, only F'_{ST}-values based on ten microsatellite loci are shown and F'_{ST}-values based on six loci are available upon request.

Mantel tests and partial Mantel tests conducted in Arlequin, using F_{ST}-values and F'_{ST}-values. Temperature difference matrices were based on the average annual air temperatures near sampling sites. All tests for correlations involving geographic distance contained only populations from the last ten years, whereas tests regarding temperature differences contained all populations with candidate gene sequences. Mantel tests and partial Mantel tests for candidate locus *TF* and selectively neutral microsatellite loci were additionally conducted on solely *D. galeata* F_{ST}- and F'_{ST}-values. All *P*-values were adjusted for multiple testing using Bonferroni correction.

Locus neutrality was tested using either a data set containing all individuals or a data set containing exclusively *D. galeata* individuals, each consisting of 11 microsatellite loci and three candidate gene markers. It was necessary to reduce the candidate marker sequence data to genotypic data, which inevitably biased the genetic distances between the haplotypes. For that reason, we performed each analysis on the raw data (called "unpooled data set") as well as on a pooled data set (Additional file C1-3: Table S15). For the pooled data set, an arbitrary threshold of 10% sequence divergence was set and haplotypes with lower values were pooled as described in Eveno *et al.* (2008).

Two outlier tests, the coalescent-based simulation method implemented in LOSITAN (Antao *et al.* 2008) and the Bayesian approach implemented in BayeScan (Foll & Gaggiotti 2008), were used on the data described above (for details see Additional file C1-1: Methods).

Acknowledgements

The authors thank N. Henning-Lucass and E. Voelker for technical assistance in sampling and molecular work, N. Brede, D. Straile, S. Hilt, P. Casper, P. Kasprzak, A. Hobaek, and A. Petrusek for providing samples, and M. Arnold and M. Pfenninger for comments prior to submission, as

well as two anonymous reviewers. This work benefits from and contributes to the Daphnia Genomics Consortium.

Additional Information

This information is available from the attached CD-ROM or from the online version of the article upon publication.

Additional file C1-1.pdf: Additional information on cloning of candidate loci and outlier tests, F'_{ST} values and haplotype networks. Figure S1. Haplotype networks, locus *M*. Figure S2. Haplotype networks, locus *EA*. Figure S3. Comparison of F'_{ST}-values at different genetic markers, all individuals. Figure S4. Comparison of F'_{ST}-values at different genetic markers, *D. galeata* only. Figure S5. Outlier tests; *D. galeata* only.

Additional file C1-2.pdf: Table S1. Taxon abundances per population. Table S2. NewHybrids results and taxonomic assignment

Additional file C1-3.pdf: Table S3. Pairwise F_{ST}-values between populations at locus *EA*. Table S4. Pairwise F_{ST}-values between populations at locus *M*. Table S5. Pairwise F_{ST}-values between populations at locus *TF*. Table S6. Pairwise neutral microsatellite F_{ST}-values between populations. Table S7. Pairwise F_{ST}-values between populations at locus *TF*; *D. galeata* only. Table S8. Pairwise neutral microsatellite F_{ST}-values between populations; *D. galeata* only. Table S9. Pairwise standardized F'_{ST}-values between populations at locus *EA*. Table S10. Pairwise standardized F'_{ST}-values between populations at locus *M*. Table S11. Pairwise standardized F'_{ST}-values between populations at locus *TF*. Table S12. Pairwise standardized neutral microsatellite F'_{ST}-values between populations. Table S13. Pairwise standardized F'_{ST}-values between populations at locus *TF*; *D. galeata* only. Table S14. Pairwise standardized neutral microsatellite F'_{ST}-values between populations; *D. galeata* only. Table S15. List of haplotype pools used for outlier tests. Table S16. Results of the Mantel tests; *D. galeata* only. Table S17. List of amino acid changing substitutions between haplotypes. Table S18. Summary of characteristics for each candidate locus. Table S19. List of all candidate genes and corresponding primers. Table S20. Primer sequences and PCR conditions for the candidate loci.

Chapter 2

A genotype-phenotype association approach to reveal thermal adaptation in *Daphnia galeata*

Maike Herrmann, Nicole Henning-Lucass, Mathilde Cordellier, and Klaus Schwenk

Abstract

Altering thermal environments impose strong selection pressures on organisms, whose local persistence depends on adaptive phenotypic plastic and genetic responses. Thus far, adaptive change is monitored using phenotypic shifts or molecular markers, although inevitable obstacles are inherent in both methods. In order to circumvent these, it is necessary to find a causal link between adaptive alleles and fitness. Combining both approaches by linking genetic analyses and life-history measurements, a potential genotype-phenotype relationship can be assessed and adaptation at the molecular level demonstrated. For our study, clonal lineages of the freshwater keystone species *D. galeata* from seven different populations distributed along a latitudinal gradient across Europe were tested for local thermal adaptation in common garden experiments. Fitness-related life-history responses were quantified under different thermal regimes and experimental clones were genotyped at three candidate gene marker loci to investigate a potential genotype-phenotype association. The analyses of the life-history data showed a significant temperature effect on several fitness-related life-history traits recorded in our experiments. However, we could not detect evidence for a direct association at neither candidate gene locus between genotypes and life-history traits. The observed phenotypic shifts might therefore not be based on the tested marker loci *EA*, *M*, and *TF*, or in general not coding sequence-based and thus rather reveal phenotypic plasticity in response to thermal variation. Nonetheless, we revealed significant genotype by environment (GxE) interactions at all tested loci, potentially reflecting a contribution of marker loci to certain life-history trait values and contribution of multiple genetic loci to phenotypic traits.

Forthcoming in the *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*.
DOI: 10.1002/jez.2070

Introduction

Anthropogenic climate change progressively alters thermal environmental conditions, which is accompanied by changes in a broad spectrum of biotic and abiotic factors, confronting our planet's biota with new selection pressures (Bellard *et al.* 2012). Local extinction of species or populations as a consequence of these changes might be avoided through either acclimatization via phenotypic plasticity or micro-evolutionary change via genetic adaptation, depending on species-specific physiological limitations and genetic architecture (Hoffmann & Sgro 2011). During the last decades, studies indicating rapid evolutionary responses to current global warming increased in number (e.g., Salamin *et al.* 2010; Geerts *et al.* 2015).

In the majority of studies, adaptive change is monitored using phenotypic shifts or molecular markers (Hansen *et al.* 2012). Phenotypic trait changes, for example obtained via long-term observations (Charmantier *et al.* 2008) or gradient analyses (De Frenne *et al.* 2013), often affect the individual's fitness directly and thus result in measurable changes of performance. However, as phenotypic traits are plastic, phenotypic change might not be based on coding sequence-based change but a result of phenotypic plastic responses. Phenotypic plasticity, such as physiological, morphological or behavioral adjustments (Whitman & Agrawal 2009) in response to environmental alterations, thus complicates the assessment of the evolutionary adaptive potential when the experimental design does not allow for a clear discrimination of the type of underlying response (Merilä & Hendry 2014).

Population genetic approaches such as genome scans for F_{ST}-outliers (Storz 2005) or landscape genetics (Manel *et al.* 2010) are alternatively used to infer loci or genomic regions potentially under selection in natural populations and to exhibit above average genetic differentiation among populations. Selective agents responsible for this genetic differentiation are subsequently deduced by establishing correlations between the distribution of genetic variability and environmental heterogeneity (recently the subject of a special issue: Jensen *et al.* 2016). Although conclusive verification of a causal relationship between adaptive alleles or genotypes and fitness determined under specific environmental conditions is necessary (Barrett & Hoekstra 2011; Hansen *et al.* 2012), the hypothesized adaptive significance of identified outlier loci could only be linked to fitness in a limited number of cases (Jensen *et al.* 2016). Genotype-phenotype relationships are still mostly examined in studies concerning human diseases and traits of agricultural importance by applying data-driven genome-wide association analyses. Now that simultaneous genotyping of high numbers of individuals and therefore genotype-phenotype association analyses have become feasible for a wide range of species through high throughput sequencing (reviewed in Korte & Farlow 2013), efforts are made to utilize this approach for ecological research and towards hypothesis-driven candidate gene association analysis (e.g., Du *et al.* 2014).

In this study, we apply a genotype-phenotype association approach to link specific genotypes and alleles of *Daphnia*, a small aquatic crustacean, with fitness measurements obtained in life-history

experiments under different temperature regimes. *Daphnia* is especially useful for population genetic approaches and experimental studies, as the parthenogenetic reproduction cycle of this genus allows the establishment of clonal lineages in the laboratory. Recent common garden experiments conducted by Henning-Lucass *et al.* (2016) showed significant effects of temperature on fitness in *Daphnia galeata*. The studied clonal lineages exhibited plastic responses to temperature in several fitness-related life-history traits (size at first reproduction, size of neonates, size of first clutch, time to first reproduction, somatic growth rate), generated by the same genotypes under various thermal conditions. Another study on the effects of temperature on gene expression revealed temperature dependent expression patterns of several genes in *Daphnia magna* (Schwerin *et al.* 2009). Three of these candidate genes, *ERNA*, *MHC-1*, and *TRY5F*, were sequenced and the trypsin coding *TRY5F* gene was found to be under selection and potentially involved in local thermal adaptation among populations of the *Daphnia longispina* species complex, comprising *Daphnia longispina*, *D. galeata*, and their interspecific hybrids (M. Herrmann, Personal Communication). The trypsin protein family plays an important role in protein digestion (von Elert *et al.* 2004) and therefore in the availability of amino acids for protein synthesis and energy supply. Since, as in many organisms, an increased rate of metabolic activity and a consequently increased energy demand is a direct consequence of elevated temperatures in *Daphnia* (Paul *et al.* 2004a), genes coding for trypsins represent a likely target for thermal selection. However, a causal link between genetic differentiation and fitness under different environmental conditions is missing.

In an effort to complement existing life-history data from *D. galeata* populations distributed along a latitudinal gradient across Europe (Henning-Lucass *et al.* 2016, Henning-Lucass in prep.) and infer the functional impact of allelic variation in candidate genes *ERNA*, *MHC-1*, and *TRY5F*, experimental clones were genotyped at temperature-relevant candidate gene markers *EA* (for *ERNA*), *M* (for *MHC-1*),and *TF* (for *TRY5F*) using amplicon sequencing. Contrary to the common approach of evaluating fitness differences among genotypes (Korte & Farlow 2013), the life-history responses under different temperature regimes in common garden experiments were quantified before genotyping experimental clonal lines at marker loci. Based on above mentioned results of M. Herrmann (Personal Communication) and Henning-Lucass *et al.* (2016), we expect to find an association between the phenotype and the genotype at *TF*. By contrast, at markers *EA* and *M*, which showed no indication for selection, we do not expect to find such an association.

Material and Methods

Life-history experiments

Sediment sampling

We sampled resting egg banks at deep water sites of six freshwater lakes along a latitudinal gradient from Northern to Southern Europe (Table C2-1). Sediment cores from the following lakes were

collected between spring 2010 and summer 2012: Ringsjön (South Sweden), Müggelsee (North Germany), Jordán reservoir (Czech Republic), Lake Constance (South Germany), Greifensee (Switzerland), and Lago di Nemi (Middle Italy). In order to avoid hatching stimuli, cores were sliced and afterwards stored in opaque sealed tubes at 4 °C. In addition, a water column sample was taken from Lake Jyväsjärvi (South Finland) in autumn 2012 (Table C2-1).

Table C2-1: Sampling sites

Lake	Location	GPS Coordinates	Altitude [m]	Sample origin
Greifensee	Switzerland	47° 21‘20“ N, 8° 40‘10“ E	435	sediment
Jordán reservoir	Czech Republic	49° 24‘55“ N, 14° 39‘49“ E	437	sediment
Lake Constance	Austria, Germany, Switzerland	47° 37‘21“ N, 9° 26‘24“ E	395	sediment
Lake Jyväsjärvi	Finland	62° 14‘22“ N, 25° 46‘34“ E	78	water column
Müggelsee	Germany	52° 26‘6“ N, 13° 38‘6“ E	32	sediment
Ringsjön	Sweden	55° 52‘1“ N, 13° 31‘25“ E	54	sediment

Hatching of dormant eggs and experimental set-up

Hatching of dormant eggs and the experimental set-up for the life-history experiments, as well as recording and calculation of life-history traits are described in detail in Henning-Lucass *et al.* (2016). Propagules were obtained through sieving of sediment samples and inoculated in jars filled with equal amounts of distilled tap water and filtered pond water. After exposition to hatching stimuli (16:8 L:D, 19 ± 1 °C, Vandekerkhove *et al.* 2005), neonates were separately transferred into jars filled with ADaM (Aachener Daphnien Medium, Klüttgen *et al.* 1994) and kept under the same conditions as the dormant propagules for parthenogenetic reproduction of clonal lines. Since *D. galeata* diapausing eggs are sexually produced, each experimental clonal line has a unique genetic background. The species status was determined through Sanger sequencing of the mitochondrial *12S* gene (Taylor *et al.* 1996) and nuclear genotyping using microsatellite loci (Thielsch *et al.* 2009). Three climate chambers were equipped with identical flow-through systems containing experimental tubes. Prior to the start of the experiment, all test animals were acclimatized to treatment conditions. Therefore, we introduced neonates from the second clutch of ‘mother individuals’ (born under standard laboratory conditions and introduced to experimental temperatures as neonates) into the flow-through system within 12 hours after birth. For each clone, four individuals were placed together in one experimental tube to ensure a sufficient number of surviving individuals per tube at the end of the experiment. During the experiments time to first reproduction [d], size at first reproduction [mm], size of first clutch [number of neonates per female], size of neonates released from the brood pouch [mm], survival rate of adult females at the time of first reproduction [%] and somatic growth rate (SGR) [mm d^{-1}] were recorded as life-history traits. The SGR was calculated using the formula $[\ln(S_t)-\ln(S_0)]\ t^{-1}$ (Lampert & Trubetskova 1996), where S_0 is the mean body size of each clonal lineage at start of the experiment and S_t the

mean body size of each clonal lineage after reaching maturity, divided by the time to first reproduction (t) in days. Experimental tubes were controlled daily for released neonates to determine the time to first reproduction. The average size of the first clutch for each tested clonal lineage was calculated by counting the unborn neonates in the carapaces in addition to the already released neonates and dividing the total amount by the number of adult females per tube. In order to capture the size of the females and neonates, we transferred all individuals of the experimental tube into ethanol after the first neonates were released from the brood pouch. The individuals of one tube were not used as independent replicates. For each clonal lineage, the average value of the four individuals per experimental tube was calculated to estimate the clonal response to each temperature regime at the respective life-history trait. Survival rate was defined as the proportion of females surviving until the release of the first clutch.

Table C2-2: Experimental design

Experiment	Temperatures [°C]	Population	Data set	# CL LH	# CL *EA*	# CL *M*	# CL *TF*
1a	15, 20, 25	Lake Constance	DS1	7	3	5	7
		Greifensee		7	6	4	6
		Total	DS2	14	9	9	13
1b	22, 25, 27	Lake Constance	DS1	6	3	5	8
1c	22, 25, 27	Lake Constance	DS1	6	3	5	8
2a	15, 20, 25	Lake Constance		6	5	6	6
		Jyväsjärvi		2	3	2	2
		Greifensee		4	4	4	4
		Jordán reservoir		7	6	6	5
		Müggelsee		6	6	2	6
		Lago di Nemi		3	2	1	1
		Ringsjön		2	2	2	0
		Total	DS3	31	28	23	24
2b	25, 27	Lake Constance		6	5	6	6
		Greifensee		4	4	4	4
		Jordán reservoir		7	6	6	5
		Müggelsee		6	6	2	6
		Lago di Nemi		3	2	1	1
		Ringsjön		2	2	2	0
		Total	DS3	28	25	21	22
Total number of clonal lines				44	43	37	42

CL LH: number of clonal lineages with quantified life-history traits; # CL *EA*, *M*, *TF*: number of clonal lineages sequenced for each locus.

Experimental design

Altogether, five separate life-history experiments were conducted between October 2011 and February 2013, using different combinations of populations and clonal lines. Test temperatures varied between experiments and covered 'natural' temperatures within the reproductive period of European *D. galeata*. They ranged from 15 °C to 25 °C, with an additional slightly increased maximum temperature set at 27 °C (Table C2-2). For the phenotype-genotype association analyses, the experimental data were grouped into three distinct data sets of clonal lines. The first set (DS1) contained only clones from Lake Constance (experiments 1a-c), the second set (DS2) contained clones from Lake Constance and Greifensee (experiment 1a), and the third set (DS3) contained clones from the seven populations Lake Constance, Greifensee, Jordán reservoir, Jyväsjärvi, Müggelsee, Lago di Nemi, and Ringsjön (experiments 2a and 2b). Thus, life-history traits for the DS1 and DS3 were recorded in several different experiments and several clonal lines are shared among different experiments within the sets (Supplementary Table C2-S1). The temperatures and clones used in each experiment are shown in detail in Table C2-2.

Genotyping

Candidate genes

Three candidate genes were chosen and found to be under selection and potentially involved in thermal adaptation in natural populations of the *Daphnia longispina* species complex (M. Herrmann, Personal Communication) prior .to this study. Schwerin *et al.* (2009) showed that they were differentially expressed between 10 °C and 20 °C in *D. pulex*. Amplification and sequencing was attempted for DNA sequences of all eleven differentially expressed proteins in this study and successful for the three loci *EA*, *M*, and *TF* within the genes *ERNA*, *MHC-1*, and *TRY5F*. Primer pairs were designed using the available DNA sequences of *Daphnia pulex* (Colbourne *et al.* 2011). *MHC-1* codes for the myosin heavy chain, a component of muscle fibers, *ERNA* codes for an endoribonuclease, a secretory enzyme involved in RNA processing and *TRY5F* codes for a trypsin, a secretory enzyme involved in protein digestion.

Amplification of candidate markers

Data on genotypic variation found in the wild was obtained from ephippia collected directly from sediment cores using a 125 µm sieve (Retsch, Haan, Germany). Ephippia were stored in 70-100% ethanol until the resting eggs were isolated. Resting eggs where extracted from the ephippial case and the DNA was prepared from single resting eggs. To genotype the clonal lines tested in life-history experiments (subsequently referred to as "experimental samples"), DNA was prepared from individual clones. DNA was extracted from both experimental clones and resting eggs from the wild using the HotShot method (Montero-Pau *et al.* 2008). We used 50 µl of lysis buffer and an equal amount of neutralization buffer. The DNA was stored at 4 °C or 20 °C. We ampli ed the three loci *M*, *EA*, and *TF* from 56 experimental samples in total. Genotyping data of 87 clonal lines

obtained directly from sediment cores (subsequently referred to as "field samples") was available. For details on PCR protocol and primers for tagged amplicon sequencing of experimental clones see Supporting Information C2-S2, Supplementary Table C2-S3; for details on PCR protocol, primers, and Sanger sequencing of field samples see Supporting Information C2-S2 and C2-S5, and Supplementary Table C2-S4. Each of the three primer pairs used for amplification binds to only one locus in the transcriptome of *D galeata*, respectively. Each of the six individual primers has, however, more than one binding site, as detected by a blast against the newly published *Daphnia galeata* transcriptome (Huylmans *et al.* 2016). Any chimeric sequences possibly generated because of this were eliminated subsequently in different steps of the bioinformatic analysis. The number of clonal lines successfully sequenced for each locus in each population is listed in Table C2-2. PCR products were barcoded using unique 5'-tagged primers specific for each clonal line. Tag sequences were of 5 bp length and differed from each other by at least two bases (Supplementary Table C2-S6). PCR products were purified using either the Agencourt AMPure Kit (Beckman Coulter, Brea, CA, USA) or the PureLink PCR Purification Kit (Invitrogen, Carlsbad, CA, USA).

Amplicon pooling

The amount of DNA in PCR amplicons was measured visually using the MassRuler Express Forward DNA Ladder Mix (Thermo Scientific, Waltham, MA, USA) for reference. Equal amounts of DNA from each individual PCR product were combined into two amplicon pools. The first, Sample 1, contained amplicons of locus *M* of ~ 400 bp sequence length and 43 ng DNA/clonal line. The other, Sample 2, contained amplicons of both loci *EA* and *TF* of ~ 630 bp and ~ 740 bp sequence length, respectively, and 23 ng DNA/clonal line.

454 targeted amplicon sequencing, quality control, and genotype reconstruction

Both pools were sequenced by the GATC company, using a Roche GS FLX sequencer and the Titanium Sequencing Kit XLR70 (Roche 454 Life Sciences, Branford, CT, USA). In a first quality control step, all reads showing any divergence from original barcode and primer sequences were discarded. The program USEARCH (Edgar 2010) and custom Python scripts were used for all subsequent procedures. Reads were filtered and truncated, whereby custom parameters (Supplementary Table C2-S7) were applied to each data set. Depending on locus and sequence orientation, the read data was sorted and merged into five globally trimmed data sets, one for locus *EA*, containing all reads, and two for each of the loci *TF* and *M*, containing either forward or reverse reads. For each data set, sequences with at least 97% sequence similarity were clustered. The true haplotype sequences in each data set were estimated using the consensus of the reads within each cluster by choosing the majority base at each site.

In order to reconstruct each clonal line's individual genotype, we mapped the quality controlled, merged data sets containing individual reads against the cluster consensus sequences. The resulting file was used to assign individual reads and their corresponding cluster consensus sequences to the

clonal lines. We thereby expected to assign one (in homozygous lines) or two (in heterozygous lines) forward and reverse cluster consensus sequences to each clonal line. However, probably due to PCR and sequencing errors, more than two cluster consensus sequences were assigned to the majority of clonal lines (to all clonal lines for loci *EA* and *TF*, and to 86.7% for locus *M*). The final reconstruction of individual genotypes was based on the support of the haplotypes in terms of read numbers. A true allele had to be supported by reads sequenced from both directions (forward and reverse) and to have a share of at least 10 % of each clonal line's reads. If more than one haplotype sequence fulfilled these criteria, the clone was assumed to be heterozygous and its genotype to be composed of the two most abundant haplotype sequences. The 10 %-threshold was set arbitrarily to balance between differences in PCR and sequencing success as source of an uneven distribution of reads among two real alleles on one hand and pseudo-alleles artificially created by PCR and sequencing errors on the other hand. For detailed descriptions of methods used for quality control, clustering, and genotype reconstruction see supplementary material (Supporting Information C2-S8 and Supplementary Figure C2-S9).

Statistical analyses

Life-history experiment

Statistical analyses of life-history data were conducted for all three data sets with IBM SPSS Statistics 22. In a first step, data were tested for normality of distribution using the Lilliefors corrected Kolmogorov-Smirnov test. Afterwards, Kruskal-Wallis (nonparametric) tests and Analysis of Variance (parametric ANOVA) were performed to compare thermal conditions. Additionally, we used general linear models (GLM) to test for the effect of temperature, population origin and the interaction of both for all life-history traits. Subsequently, post-hoc tests (Fisher's least significant difference, LSD) were performed to compare measured traits between two temperature treatments.

Association analyses

The programs PLINK v.1.07 (Purcell *et al.* 2007) and mvBimBam (Stephens 2013) were used to test for association between genotype and phenotype under different temperatures. PLINK uses univariate linear regression to test for association between each SNP and each phenotypic trait individually, whereas mvBimBam uses a multivariate approach to test for a relationship between the phenotypic trait and all SNPs at once. Only non-redundant SNPs were included into the analyses.

Association testing with Plink was performed using the "qassoc" command. Settings were applied to correct for testing of multiple SNPs and for population stratification in the data sets by permuting within populations. Additionally, clones with more than 10% missing genotypes and SNPs with a minor allele frequency below 1 % were excluded prior to testing (options: "--within --assoc --all-pheno --mperm 1000 --mind 0.1 --geno 0.1 --maf 0.01"). All *P*-values were corrected for

multiple testing using the false discovery rate (FDR) as implemented in the R package q-value v1.38.0 (Dabney & Storey 2003).

For multivariate association testing with mvBimBam, traits with values for less than three clonal lines and clonal lines with missing genotype or missing values for any of the phenotypic traits were excluded from the analyses. Therefore, all analyses were conducted for each of the loci *EA*, *M*, and *TF* separately and included varying numbers of clonal lines and phenotypic traits (Supplementary Table C2-S10). Prior to analyzing, phenotypic values in DS2 and DS3 were corrected to account for possible population stratification using linear regression. The covariate-corrected phenotypic values (i.e., the residuals) were used for further analysis. For all data sets, phenotypic values were quantile transformed to a standard normal distribution and tested for multivariate normality (MVN) using Royston's multivariate normality test as implemented in the R package MVN v3.7 (Korkmaz *et al.*). The only data set tested positive for MVN, DS3, was subsequently tested in mvBimBam (using the "mph 2" option and default settings). Genotypes were tested against somatic growth rate, size of the first clutch, and neonates size at 15 °C, 20 °C, 25 °C (experiment 2a), 25 °C and 27 °C (experiment 2b) in one analysis, and against somatic growth rate, time to maturity, and neonates size at the same temperatures in a separate analysis.

PLINK was further used to test for differences in genotype-phenotype-association between two environments (genotype environment interaction, GxE) by comparing the difference between the regression coefficients of each association. GxE between genotypes and temperature was tested for somatic growth rate, size of the first clutch, and size of neonates for DS1, DS2, and DS3. All possible pairs of temperature regimes within experiments conducted simultaneously were analyzed (Supplementary Table C2-S11). The phenotypic values of DS2 and DS3 were corrected for containing multiple populations using linear regression as described above. All analyses were conducted for each of the loci *EA*, *M*, and *TF* separately, using the command "gxe", filtering options as described above and temperature set as the covariate. Clonal lines with missing genotype or phenotype values for any of the phenotypic traits were excluded from the analyses. All *P*-values were corrected for testing multiple testing using the false discovery rate (FDR) as described above.

Population differentiation and genetic diversity in field and experimental samples

The clones tested in life-history experiments have possibly experienced a loss of genetic diversity at our loci of interest due to the sampling bias, uneven hatching success under laboratory conditions, subsequent laboratory selection, and the strategy we applied to obtain the sequences. In order to test for loss of genetic diversity in experimental clones, we compared the pairwise genetic differentiation among the populations in experimental and field samples. Due to the applied clustering strategy, sequences with less than 3 % differences could not be distinguished in experimental clones. Thus, the Sanger-sequenced haplotypes of the field samples were also clustered, using USEARCH and applying the same strategy as described for the amplicon sequencing data (Supporting Information C2-S8 and Supplementary Figure C2-S9).

Pairwise genetic differentiation among populations with both experimental and field samples (*EA* and *TF*: Greifensee, Jordán reservoir, Müggelsee, and Lago di Nemi; *M*: Greifensee, Jordán reservoir, and Lago di Nemi) was estimated for each locus and each data set (unclustered field samples, clustered field samples, and experimental samples). F_{ST}-values were computed using genetic distances between haplotypes based on AMOVAs and the Kimura 2 parameter (K2P) model as implemented in Arlequin ver. 3.5.1.2 (Excoffier & Lischer 2010). F_{ST} *P*-values were computed using 10000 permutations and corrected for multiple testing using the Bonferroni method.

We also exemplarily compared the haplotypic diversity of experimental samples at locus *TF* to the diversity found in the wild. In order to find haplotypes in the field samples corresponding to the experimental haplotypes despite the clustering process, we mapped the clustered field *TF* haplotype sequences to the experimental *TF* haplotype sequences. Both haplotype sets were trimmed to equal lengths beforehand and otherwise processed using the same strategy as described for the experimental samples (Supporting Information C2-S8). Fisher's exact tests were used to test for independence between observed haplotype frequencies and sample origin in all populations with both experimental and field samples.

Results

Life-history experiments

Analyses of life-history data showed that temperature significantly affected several fitness-related life-history traits in our samples. In data set DS2, temperature had a significant effect on time to first reproduction (GLM: F (2,31) = 25.582, P < 0.001) and somatic growth rate (GLM: F (2,31) = 17.36, P < 0.001), revealing that higher temperatures lead to an accelerated growth and faster reproduction in clones from Lake Constance and Greifensee (Figure C2-1A and B). In DS1, temperature had the same effect on time to first reproduction (experiment 1b: Kruskal-Wallis-test: ^{2} (2, N = 74) = 12.16, P = 0.002, experiment 1c: ANOVA: F (2,71) = 22.60, P < 0.001) and somatic growth rate (experiment 1b: ANOVA: F (2,71) = 9.61, P = 0.001, experiment 1c: ANOVA: F (2,71) = 11.34, P = 0.001). However, life-history trait values differed between experiment 1c and 1b (Figure C2-2), despite identical experimental set-ups with same temperature regimes and an almost identical set of clones from Lake Constance (one clonal line differed between 1b and 1c at 22 °C for locus *M* and another at 22 °C and 25 °C for locus *TF*, of which none had a unique genotype at these loci). Whereas in experiment 1b, clonal lineages of the 27 °C treatment differed significantly in time to first reproduction and somatic growth rate from clones in 22 °C and 25 °C, in experiment 1c the clonal lineages of the 22°C treatment reproduced and grew significantly faster than those in 25 °C and 27 °C. Additionally, temperature affected adult female size (experiment 1c: ANOVA: F (2,71) = 8.67, P = 0.003) as well as neonate body size (experiment 1c: Kruskal-Wallis-test: ^{2} (2, N = 74) = 8.64, P = 0.013) significantly in experiment 1c, with smallest individuals

collected from the 27 °C treatment. In DS3, GLM analyses and pairwise comparisons of 15 °C, 20 °C and 25 °C treatments via post-hoc LSD tests revealed a significant effect of temperature on all six tested life-history traits (Table C2-3). Furthermore, inter-population differences were found for size of neonates, size of first clutch, time to first reproduction, survival rate, and somatic growth rate (Table C2-3), which indicates local adaptation. Comparing 25 °C and 27 °C in DS3, no temperature effect between the two high temperature treatments was found, but instead inter-population differences for female size (GLM: F (5,47) = 3.638, P = 0.008), time to first reproduction (GLM: F (5,47) = 2.896, P = 0.025), somatic growth rate (GLM: F (5,47) = 3.003, P = 0.021), and survival rate (GLM: F (5,50) = 5.919, P < 0.001).

Genotype-phenotype association

We found no evidence for a direct association between genotypes (for each candidate gene locus) and any of the life-history traits in the association analyses, as exemplified for time to first reproduction and somatic growth rate in Figures C2-1C to H. For all three candidate genes was observed that clones frequently exhibited similar phenotypic responses although they carried different haplotypes. We also found a wide variation in phenotypic trait values among clones carrying identical candidate gene haplotypes. Similar results were observed for multivariate phenotypic responses: None of the univariate association analyses conduced in PLINK revealed a significant correlation and all $\log_{10}$ Bayes Factors obtained from the multivariate association analysis with mvBimBam on DS3 were below 0.5. Even with a prior odd as low as one (i.e., under the assumption that the presence and absence of an association are equally likely), no significant correlation between genotype and multivariate phenotype could be detected.

Significant GxE interactions across two temperature regimes were found for all tested loci (Table C2-4). Significant correlations between the genotype at locus *EA* and mean size of first clutch were revealed for temperature pair 15 °C/20 °C in experiment 1a of DS1 (including clones only from Lake Constance) and temperature pair 15 °C/25 °C in experiment 1a of DS2 (including clones from Lake Constance and Greifensee). The genotype at locus *M* significantly correlated with mean somatic growth rate (for temperature pair 15 °C/25 °C in experiment 1a of DS2) and mean size of the first clutch (for temperature pair 22 °C/27 °C in experiment 1c of DS1, and temperature pairs 22 °C/25 °C and 15 °C/25 °C in experiment 1a of DS2). Mean size of neonates was also correlated with genotype at locus *M* for temperature pair 15 °C/25 °C in both experiment 1a of DS2 and experiment 2a of DS3 (which was the only significant result for the data set containing all populations). Significant associations between the genotype at locus *TF* were revealed for mean somatic growth rate (for temperature pair 25 °C/27 °C) and mean size of first clutch (for temperature pair 22 °C/27 °C) both in experiment 1c of DS 1.

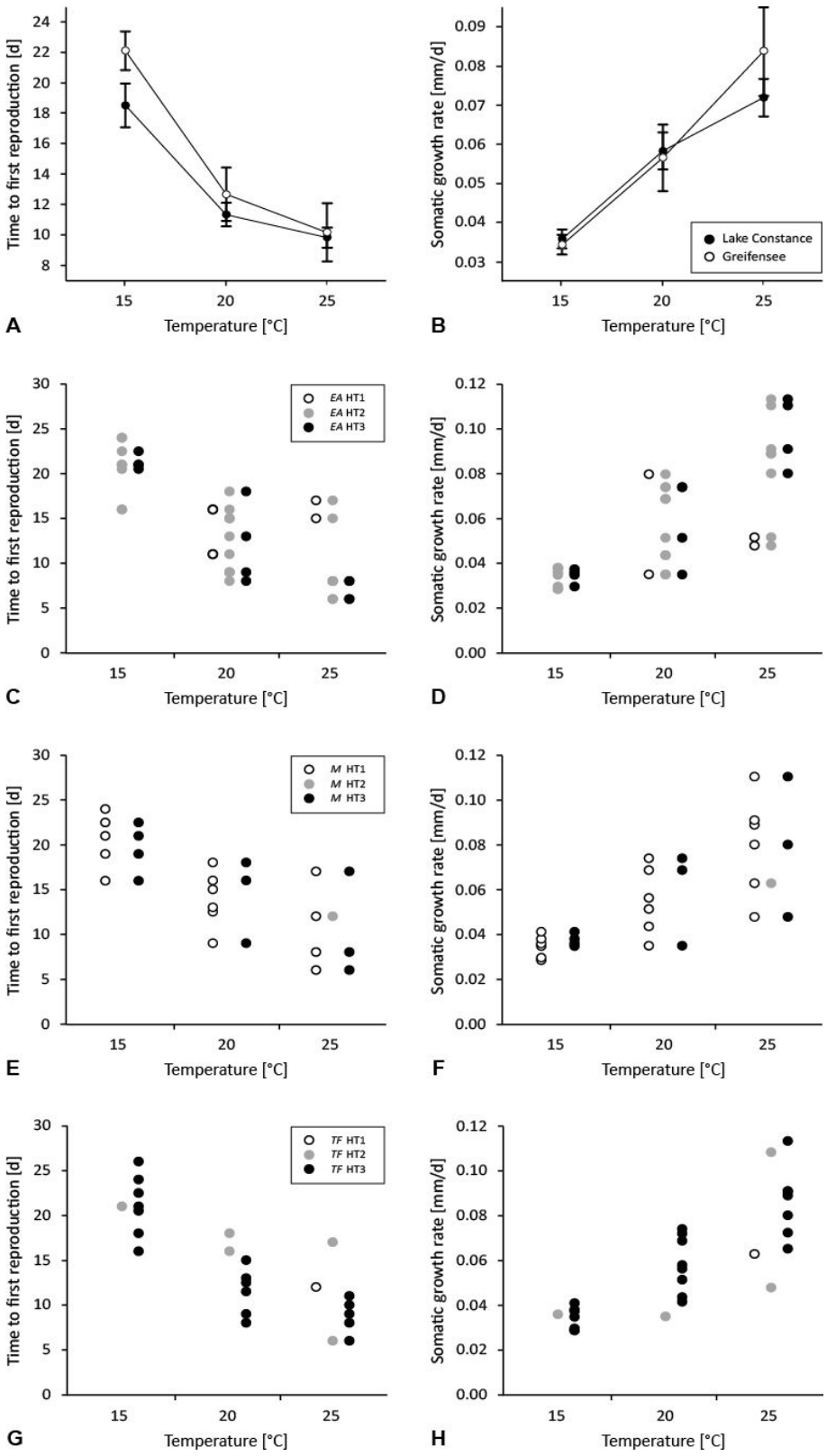

Figure C2-1: Population reaction norms and haplotype frequencies for DS2. Experimental samples from Lake Constance and Greifensee showing a significant effect of the thermal regime on time to first reproduction (A) and somatic growth rate (B). In both populations, increased temperatures resulted in an earlier reproduction and accelerated growth. Given are means for life-history trait values. Haplotypes at the candidate loci show no association with phenotypic response to different temperatures in both time to first reproduction (*EA*: C, *M*: E, and *TF*: G) and somatic growth rate (*EA*: D, *M*: F, and *TF*: H)

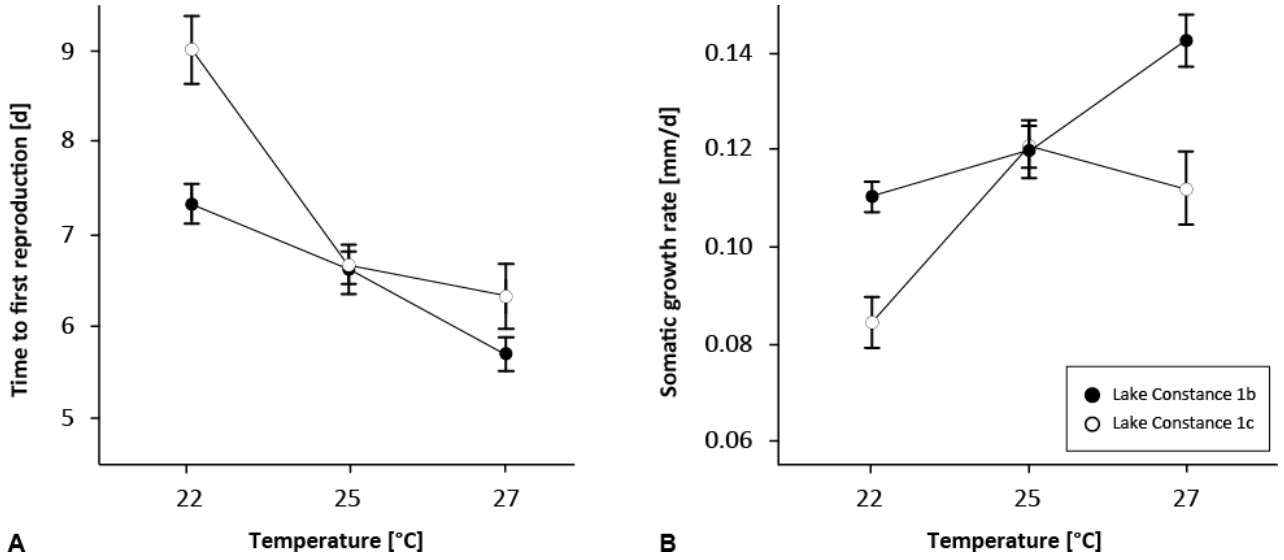

Figure C2-2: Differences in life-history trait values among almost identical sets of clones from Lake Constance (experiment 1b and 1c of data set DS1) for time to first reproduction and somatic growth rate. Given are means ± SE for both experiments

Table C2-3: Life-history experiment

Life-history trait	*N*	Ind. variable	df	*F*	*P*	*P* (LSD)		
						15 vs. 20 °C	20 vs. 25°C	15 vs. 25 °C
Size at first reproduction [mm]	85	Temperature	2	17.747	**< 0.001**	**< 0.001**	0.284	**< 0.001**
		Population	6	1.007	0.428			
		T x P	12	1.285	0.253			
Size neonates [mm]	85	Temperature	2	5.309	**0.007**	**0.033**	**0.018**	**< 0.001**
		Population	6	4.220	**0.001**			
		T x P	12	1.045	0.419			
Size of first clutch [neonates/female]	92	Temperature	2	9.987	**< 0.001**	**< 0.001**	0.548	**< 0.001**
		Population	6	6.033	**< 0.001**			
		T x P	12	3.109	**0.001**			
Time to first reproduction [d]	85	Temperature	2	94.908	**< 0.001**	**< 0.001**	**0.023**	**< 0.001**
		Population	6	3.760	**0.003**			
		T x P	12	0.937	0.512			
Survival rate [%]	92	Temperature	2	3.829	**0.026**	**0.035**	0.213	**0.001**
		Population	6	12.824	**< 0.001**			
		T x P	12	1.052	0.413			
Somatic growth rate [mm/d]	85	Temperature	2	51.198	**< 0.001**	**< 0.001**	**0.024**	**< 0.001**
		Population	6	3.217	**0.008**			
		T x P	11	0.759	0.679			

General linear model testing for the effect of temperature, inter-populational differences, and the interaction of both (T x P) comparing 15 °C, 20 °C, and 25 °C in experiment 2a (DS3). Post-hoc tests (Fisher's least significant difference, LSD) additionally testing for differences in-between temperature regimes. Variation in sample size (*N*) are a result of clonal lines not reaching the end of the experiment. Ind. Variable: independent variable. Significant *P*-values (< 0.05) are highlighted in bold.

It is important to note, however, that the number of possible GxE tests in DS1 was reduced because of lack of observations for both genotypic and phenotypic values in both temperature regimes (this concerned loci *EA* and *M*) and monomorphic genotypes (this concerned locus *TF*). The locus most affected was *EA*, which was involved in only one test on DS1.

Table C2-4: GxE interactions across two temperature regimes for mean somatic growth rate (SGR), mean size of first clutch (cl1S), and mean size of neonates (neoS)

Data set		Test	*P*-value SGR			*P*-value cl1S			*P*-value neoS		
			EA	*M*	*TF*	*EA*	*M*	*TF*	*EA*	*M*	*TF*
DS1	1a	15 °C x 20 °C	ns	ns	NA[b]	**< 0.001**	ns	NA[b]	ns	ns	NA[b]
		20 °C x 25 °C	NA[a]	NA[a]	NA[b]	NA[a]	NA[a]	NA[b]	NA[a]	NA[a]	NA[b]
		15 °C x 25 °C	NA[a]	NA[a]	NA[b]	NA[a]	NA[a]	NA[b]	NA[a]	NA[a]	NA[b]
	1b	22 °C x 25 °C	NA[a]	ns	ns	NA[a]	ns	ns	NA[a]	ns	ns
		25 °C x 27 °C	NA[a]	ns	ns	NA[a]	ns	ns	NA[a]	ns	ns
		22 °C x 27 °C	NA[a]	ns	ns	NA[a]	ns	ns	NA[a]	ns	ns
	1c	22 °C x 25 °C	NA[a]	ns	ns	NA[a]	ns	ns	NA[a]	ns	ns
		25 °C x 27 °C	NA[a]	ns	**0.009**	NA[a]	ns	ns	NA[a]	ns	ns
		22 °C x 27 °C	NA[a]	ns	ns	NA[a]	**< 0.001**	**0.045**	NA[a]	ns	ns
DS2	1a	15 °C x 20 °C	ns	ns	ns	ns	ns	ns	ns	ns	ns
		20 °C x 25 °C	ns	ns	ns	ns	**0.03**	ns	ns	ns	ns
		15 °C x 25 °C	ns	**0.03**	ns	**0.01**	**0.03**	ns	ns	**0.03**	ns
DS3	2a	15 °C x 20 °C	ns	ns	ns	ns	ns	ns	ns	ns	ns
		20 °C x 25 °C	ns	ns	ns	ns	ns	ns	ns	ns	ns
	2b	15 °C x 25 °C	ns	ns	ns	ns	ns	ns	ns	**0.04**	ns
		25 °C x 27 °C	ns	ns	ns	ns	ns	ns	ns	ns	ns

Conducted in PLINK. Significant *P*-values (< 0.05) are highlighted in bold.

[a] < three observations in at least one of the two temperatures

[b] monomorphic genotypes (minor allele frequency < 0.01)

Diversity in field and experimental populations

Mapping *TF*-haplotypes from field samples to experimental sample reads in USEARCH allowed us to find corresponding haplotypes in samples of field and experimental origin. Clustering field sample haplotypes exhibiting less than 3% divergence among each other resulted in a perfect correspondence of *TF*-haplotypes from field and experimental samples. For each *TF*-haplotype found in the experimental samples, a corresponding haplotype was found in the field samples, but not vice-versa (Figure C2-3, Supplementary Table C2-S11).

Over all clonal lines, unclustered field samples contained more haplotypes than experimental samples and still contained a higher number of haplotypes after clustering (Figure C2-4). Because of the higher sample size in the field samples, none of these differences in haplotype count are significant. Samples from individual populations of both field and experimental origin contained

similar numbers of different haplotypes but showed vast differences in their frequencies in some of the populations. This observation was confirmed by the results of the Fisher's exact tests of independency, which show dependency of genotype frequencies and sample origin in the populations of Jordán reservoir and Lago di Nemi, i.e., in two of five tests (Figure C2-3, Supplementary Tables C2-S12 and C2-S13). Pairwise F_{ST}-values between unclustered field samples were between 0.193 (Greifensee versus Müggelsee) and 0.421 (Jordán reservoir versus Müggelsee) at locus *EA*, between 0.219 (Greifensee versus Lago di Nemi) and 0.419 (Jordán reservoir versus Lago di Nemi) at locus *M*, and between zero (Greifensee versus Müggelsee) and 0.766 (Müggelsee versus Lago di Nemi) at locus *TF* (Supplementary Table C2-S14). Clustering haplotype sequences with less than 3% difference increased F_{ST}-values in all but two comparisons (Greifensee versus Müggelsee & Greifensee versus Lago di Nemi) at locus *EA* (by 2% and 37%, respectively) but decreased all F_{ST}-values but one (no change between Greifensee and Lago di Nemi) by 8% and 48% at locus *M*. At locus *TF*, F_{ST}-values both increased and decreased by notably lower rates between 0 to 10%. Regardless of the locus, 93% of the pairwise population comparisons conducted on the unclustered field sample data set and 80% of the comparisons conducted on the clustered sample data sets yielded results significantly different from zero, while none of the F_{ST}-values between any experimental population pair differed significantly from zero.

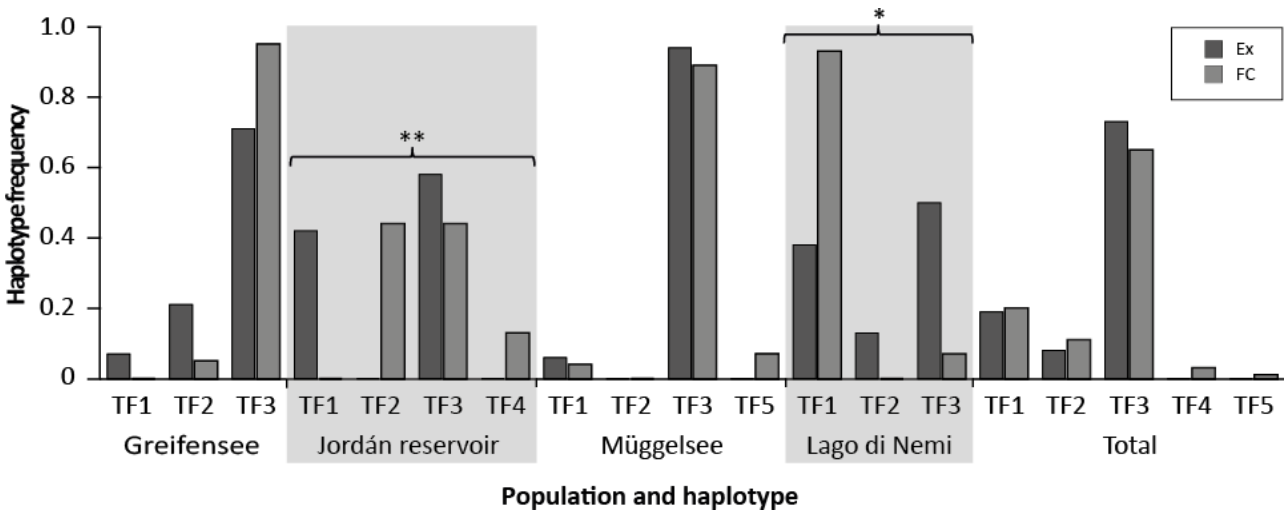

Figure C2-3: Distribution of haplotypes among populations at locus *TF*. Asterisks indicate dependency between observed haplotype frequencies and sample origin (Fisher's exact test; * $P < 0.01$; ** $P < 0.001$). FC: clustered field samples, Ex: experimental samples

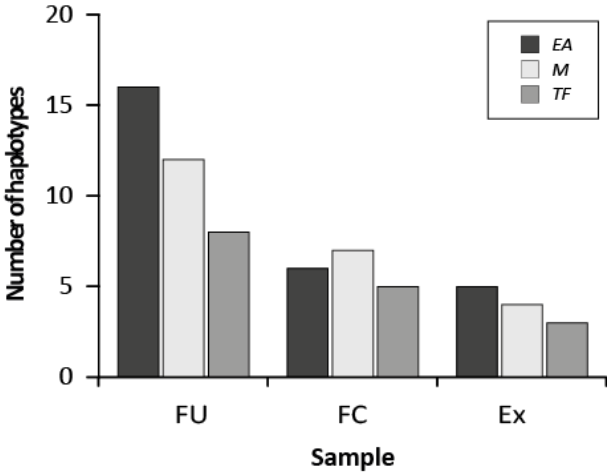

Figure C2-4: Distribution of haplotypes at all candidate loci among the different samples. FU: unclustered field samples, FC: clustered field samples, Ex: experimental samples

Discussion

In order to estimate the adaptive potential of *D. galeata*, we investigated the genotype-phenotype relationship by comparing variation in (putatively temperature-relevant) genetic markers and life-history traits under different thermal conditions. A significant effect of temperature on life-history traits (size at first reproduction, size of neonates, size of first clutch, time to first reproduction, and somatic growth rate) as well as indications for micro-evolutionary responses in *D. galeata* have previously been shown by Henning-Lucass *et al.* (2016). The findings of the authors are in accordance with a wealth of studies showing temperature effects on fitness-related life-history traits in *Daphnia* (e.g., Rinke & Petzoldt 2003; Bernot *et al.* 2006). Since one of the three marker loci used in our study, *TF*, was recently found to be under natural selection in the *D. longispina* species complex (M. Herrmann, Personal Communication), we expected to find an association between the genotype at marker *TF* and the phenotype assessed in previous thermal adaptation studies.

In contrast to our expectations, no evidence for a direct genotype-phenotype association could be detected at any of the three candidate gene loci. Additionally, and in contrast to our hypothesis, we revealed significant genotype by environment interactions not only at locus *TF* but at all three tested loci, possibly reflecting a contribution of the marker loci to certain life-history trait values. Whereas significant GxE interactions with the genotype at locus *TF* were found only in DS1, including solely clonal lineages from Lake Constance, interactions with the genotype at locus *EA* were also found in DS2, including clonal lineages from Lake Constance and Greifensee. For the genotype at locus *M,* significant GxE interactions were revealed for DS1, DS2, and even DS3, including all tested populations. The inconsistency of GxE results in DS1 between experiments 1b

and 1c, conducted under the same temperature treatments, are most possibly a result of the above mentioned differences in phenotypic trait values between those two experiments (Figure C2-2). Such variations in measurable phenotypic traits might reflect fluctuations in experimental conditions beyond our control, e.g., temporal variations in water and nutrition quality.

Significant GxE interactions might indicate the contribution of multiple genetic loci to the phenotypic traits, as the lack of direct genotype-phenotype associations shows that the genotype is not determining the absolute values of the phenotypic traits, but rather the amount by which a phenotypic trait differs among two temperature regimes. If true, other loci contributing to the multilocus phenotypes would outweigh the candidate genes' contribution to the absolute trait values and thus obscure direct associations. Since the marker loci are not significantly contributing to the absolute trait value, all marker loci are, against our initial assumption, unlikely to be involved in thermal selection acting on the life-history-traits analyzed in this study. Instead, the previously observed correlation between population differentiation at locus *TF* and environmental temperature was possibly caused by a third factor associated with both parameters, such as cyanobacterial blooms (Gademann & Portmann 2008; Paerl & Huisman 2008) or oxygen supply. The experimental basins of the life-history experiment were provided with an additional oxygen supply, in order to minimize secondary effects of temperature differences and observe the actual temperature effect disentangled from the amount of dissolved oxygen. This might however have weakened the effect of the thermal regime on life-history traits, as the aerobic performance has been shown to set the capacity of the lower and upper thermal limits (Pörtner 2001). In order to verify these hypotheses, a follow-up study examining the genotype-phenotype relationship under varying feeding regimes with different concentrations of cyanobacteria and under different oxygen concentrations would be necessary.

However, our results do not rule out an involvement of the candidate genes in thermal adaptation. Artificial selection experiments on *Drosophila* showed that thermal adaptation can be achieved through different mechanisms, depending on the specific selection pressure, and result in highly varying responses (Sørensen *et al.* 2007). As the phenotypic traits analyzed here are not indicative for all possible mechanisms of thermal adaptation, potential phenotype-genotype associations (for example during acute stress) would have been inevitably missed. Observing genotype-phenotype relationships in this context requires exposure to these specific conditions and possibly quantifying more specific traits as, e.g., the critical thermal maximum, CT_{max}. This measure allowed Geerts *et al.* (2015) to demonstrate the evolution of increased tolerance to higher temperatures in *D. magna*. Yampolsky *et al.* (2014a) used the time until immobilization (T_{imm}) as a measurable trait to show adaptive phenotypic plasticity in temperature tolerance and potential local thermal adaptation in *D. magna*. Similar measures were used in a genotype-phenotype association study, revealing that the genotype at the phosphoglucose isomerase (*PGI*) locus significantly affected several life-history traits and chill-coma recovery time resistance in the copper butterfly *Lycaena tityrus* (Karl *et al.* 2008).

Alternative explanations for the absence of direct genotype-phenotype associations might originate from limitations of the bioinformatic analyses, the experimental set-up, and laboratory selection of genotypes which might have occurred during the establishment of clonal lineages prior to the experiments. For example, the read mapping left us with a surprisingly high number of clusters assigned to the individual clonal lines and our subsequent clustering strategy potentially reduced the variability of observed genotypes.

Furthermore, the potential exclusion of less fecund clonal lineages from the experiment, due to an artificially set fecundity threshold of four neonates per tube, might have lowered the sample size and introduced a non-random bias in the range of tested genotypes, reducing the variability of the phenotypic responses. In this context, the comparison of genotype diversity between natural (field) samples and laboratory (experimental) samples at locus *TF* revealed a dependence on origin of the genotype in two of the tests. Additionally, genetic differentiation among populations was non-existent in laboratory samples across all loci (contrasting with field samples). The comparison of pairwise F_{ST}-values among the field populations (clustered and unclustered haplotypes) shows that the difference was not induced by methodological issues (sequencing technology and bioinformatic analyses). Hence, these results show that the analyzed clonal lines do not represent the full spectrum of variance found in field populations and therefore might imply the occurrence of laboratory selection during the process of hatching and/or rearing of the clonal lines intended for the experiments. Unfortunately, this bias is further enhanced by the limited sample size included in the association tests, potentially concealing locus *TF's* contribution to the phenotype under different conditions. Gómez and Carvalho (2000) observed similar patterns in rotifers when they compared the diversity of pelagic samples with unhatched resting eggs based on microsatellite markers. The authors suggested an uneven hatching success and/or selection after hatching as a potential cause for the differences they found among both groups. However, studies on *D. mendotae*, using allozyme locus *PGI* (Kerfoot *et al.* 2004), as well as on *D. galeata*, using allozyme loci *PGI* and *PGM* (Jankowski & Straile 2003; Brede 2008), showed matching genotype frequencies among pelagic samples and samples composed of hatched individuals from sediment of corresponding periods. These findings demonstrate that the effect of laboratory culturing on allele frequencies can differ, either because of laboratory selection or the sampling-induced bottleneck, depending on the locus of interest. This should be considered to ensure a satisfying degree of genetic diversity to cover a representative spectrum of the natural variation and study evolutionary processes.

However, applied GLM analyses of life-history data showed a general temperature effect and inter-population differences in measured traits. Thus, our sample size and the established genetic variability were sufficient for demonstrating phenotypic variation in observed traits and populations. Therefore, the absence of a genotype-phenotype association might either suggest that the phenotypic shifts in measured life-history traits are genetically-based but not directly on the tested marker loci *EA*, *M*, and *TF*, or it might reveal the lack of an underlying micro-evolutionary

thermal response. Thus, observed phenotypic variation between temperature regimes would indeed be either a non-adaptive stress response (Teplitsky *et al.* 2008) or a plastic response in terms of adaptive thermal acclimation (Charmantier *et al.* 2008). If the latter is the case, adaptive genetic divergence among phenotypes may be located in regulatory elements controlling candidate gene expression levels. An analysis of the relationship between phenotypic responses and candidate gene expression levels may thus clarify their role in thermal adaptation. In the past, plasticity has already been shown to contribute more strongly to phenotypic shifts than micro-evolutionary changes, although phenotypic differences were initially often assumed to have a genetic basis (Hoffmann & Sgro 2011; Merilä & Hendry 2014). Comparing populations along a latitudinal gradient in DS3, the inter-population differences in phenotypic traits suggested adaptation to locally prevailing temperature regimes. However, these differences do not allow for an explicit assumption of local thermal adaptation and might also be based on adaptation to local factors other than temperature, a problem inherent to gradient studies (De Frenne *et al.* 2013).

Conclusion

Although we could not find a causal relationship between genetic variation at candidate loci and life-history phenotypes and suggest phenotypic plasticity as basis for phenotypic shifts in life-history traits, a potential association of candidate genes and thermal adaptation should not completely ruled out. Whether the lower diversity of laboratory cultured clones versus field populations prohibited the detection of associations remains unresolved. In order to estimate the ecological relevance of genotype-phenotype interactions, it is essential to ensure that, in the laboratory samples, the genotype frequencies at the loci of interest match the frequencies occurring in field populations. This can be warranted by either using a high sample size or, should that not be possible, choosing the samples intended for experiments according to their genotypes. In order to avoid obscuring of associations by different genetic backgrounds among the samples, Rellstab *et al.* (2015) also recommend maximizing the fraction of genome used in analyses. Altogether, examining whether phenotypic and genotypic variation coincide is inevitable to demonstrate adaptation at the molecular level. For that reason, instead of focusing on individual aspects, methodological frameworks which incorporate genotypic as well as phenotypic information, including expression profiles and environmental factors are necessary and should be applied in order to reveal the genetic basis underlying adaptive variation.

Acknowledgements

We thank N. Brede, S. Hilt, P. Spaak, M. Moest, and D. Straile for providing sediment cores, K. Pulkinnen for providing clones, P. Kreuzer, E. Völker, and E. N. Sirait for technical assistance in sampling and molecular work, as well as E. M. Wörner, B. Kaltwasser, and S. Wech for technical support. The present study was conducted at the Biodiversity and Climate Research Centre (BiK-F), Frankfurt a. M. and supported by the research funding program "LOEWE-Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of Hesse's Ministry of Higher Education, Research, and the Arts.

Supporting Information

This information is available from the attached CD-ROM.

Additional Supporting Information may be found online in the supporting information tab for this article.

Chapter 3

Population transcriptomics in *Daphnia*: the role of thermal selection

Maike Herrmann, Suda Parimala Ravindran, Klaus Schwenk, and Mathilde Cordellier

Abstract

The complex interplay of forces determining genetic divergence among populations complicates the discovery of the genetic basis underlying local adaptation. Here, we utilized a combined reverse ecology and population transcriptomic approach to assess the contribution of thermal selection to population differentiation, thereby considering transcriptome-wide variation in both gene expression profiles and DNA sequences. We compared transcriptomes among four *Daphnia galeata* populations and identified transcripts potentially responding to local thermal selection. Based on an extensive literature search, a comprehensive database of candidate genes possibly under thermal selection in arthropods was composed to identify orthologous transcripts in *D. galeata*.

Although a large amount of transcripts which may contribute to local thermal adaptation was identified, temperature-relevant candidate genes were not overrepresented compared to the global gene set; suggesting that thermal selection might play a minor role in divergence among *Daphnia* populations. However, overrepresentation of temperature-relevant candidate genes among transcripts strongly contributing to sequence divergence among two populations (Greifensee and Müggelsee) indicated local thermal selection acting on the coding sequence level. Furthermore, we found that the majority of genes contributing strongly to sequence divergence did not contribute strongly to divergence at the expression level and vice versa. However, the affected gene functions were largely consistent between the two data sets, suggesting that genetic and regulatory variation constitute alternative routes for responses to natural selection. Our combined utilization of a population transcriptomics approach and literature based identification of ecologically informative candidate genes represents a useful methodology with a wide range of applications in evolutionary biology.

Submitted to *Molecular Ecology*

Introduction

Genetic divergence among populations is shaped by a complex interplay of forces, including the populations' phylogeographic history, gene flow, random genetic drift, and divergent selection (Nosil *et al.* 2009). Varying degrees and directions of selective pressures, and the genetic architecture of traits under selection, combined with signals reflecting evolutionary neutral processes, make it challenging to unravel the genetic basis of phenotypic traits under selection and thus to detect local adaptation to a particular environmental factor in natural populations. We therefore utilized a reverse ecology approach (Li *et al.* 2008) that allows us to assess the contribution of particular stressors or influences in shaping expression and sequence divergence among natural populations. For this purpose, we take advantage of the vast and growing body of literature on gene functions and on candidate genes for various environmental factors. These resources can be used to compile sets of candidate genes relevant for any particular environmental factor. The combination of such ecologically informative candidate gene sets with methods to identify loci strongly contributing to population divergence within transcriptome- or genome-wide SNP data allows for assessing any environmental factor's importance in shaping the divergence among populations. Furthermore, it allows for the identification of the involved gene functions. Importantly, this approach can be utilized not only for a variety of environmental factors but is also helpful in identifying loci of ecological and evolutionary interest in a wide variety of taxa, especially in non-model organisms, where information on the relationship between genes and phenotypes is scarce.

Assessing the genetic basis underlying local adaptation is essential to identify the constraints determining the geographic and ecological ranges of a species and ultimately to evaluate their adaptive potential (Merilä & Hendry 2014). Phenotypic variation observed among populations alone is insufficient for this aim since it is not necessarily immediately adaptive but often merely plastic (James *et al.* 1997; Ayrinhac *et al.* 2004; Mason & Taylor 2015). Yet, phenotypic plasticity may be under selection and contribute to local adaptation (Levine *et al.* 2011; Harrison *et al.* 2012). A fundamental mechanism underlying phenotypic plasticity is gene expression regulation (Wray *et al.* 2003). Differentiation of gene expression might reflect directional selection as a consequence of ecological influences (Romero *et al.* 2012; Xu *et al.* 2015) and was suggested to be specifically relevant in early stages of divergence among populations (Toews *et al.* 2016). Since selection acts on the expression as well as on the sequence level, transcriptome data are particularly useful in revealing genes contributing to phenotypic plasticity and patterns of local adaptation.

Temperature is an important selective agent that severely affects organisms through its impact on physical and chemical processes. Ectotherms possess a range of physiological and behavioral responses to cope with the impact of temperature (Gibert & De Jong 2001). In *Daphnia*, a classic model for studying on phenotypic plasticity, examples of phenotypic traits affected by temperature include life-history processes, metabolism, and gene expression (reviewed in Wojtal-Frankiewicz

2012). Thus, *Daphnia* has been subject of several population genetic studies on climate change. Temperature adaptation and adaptive phenotypic change (of genotype composition, life-history traits, temperature tolerance, and gene expression levels) in response to temperature increase has been shown in selection experiments as well as over time in natural populations (Van Doorslaer *et al.* 2009b; Van Doorslaer *et al.* 2010; Geerts *et al.* 2015; Henning-Lucass *et al.* 2016; Jansen *et al.* 2017). Recent studies suggest local thermal adaptation in *Daphnia pulex* and *Daphnia magna* (Williams *et al.* 2012; Yampolsky *et al.* 2014a), and enhanced tolerance to elevated temperatures in *D. magna* from warmer climates due to adaptive plasticity in hemoglobin expression (Yampolsky *et al.* 2014a).

Verifying local adaptation of phenotypic divergent traits or pinpointing the genetic basis of locally adapted traits in *Daphnia*, however, has been mostly inconclusive so far (e.g., Williams *et al.* 2012; Herrmann *et al.* in press). The detection of local thermal adaptation and underlying genes is possibly prevailed by other factors that put strong selective pressure on phenotypic traits and disguise the signal of thermal selection, such as nutrient supply (e.g., Orcutt & Porter 1983; Spaak *et al.* 2012) and parasite (e.g., Decaestecker *et al.* 2007) or predator pressure (e.g., Cousyn *et al.* 2001). A promising approach to overcome this difficulty and reveal candidate genes likely underlying thermally selected traits was employed by Yampolsky *et al.* (2014b), who focused on differences among heat-tolerant and heat-sensitive *D. magna* clones in response to optimal and severely increased "substressful" temperatures. The contribution of these genes to population divergence and local adaptation, however, remains unclear since *Daphnia* is well known for high amounts of clonal variation observed within and between populations (Mitchell & Lampert 2000). Another prominent approach employed in the majority of reference studies (for temperature-relevant candidate genes) is to assess the transcriptomic response to acute thermal stress (e.g., Jansen *et al.* 2017). While these studies certainly reveal genes affected by thermal selection, they tell little about the role of these loci in moderate conditions or about loci involved in local thermal adaptation but not directly affected by acute thermal stress. Yet, differences in performance among populations in intermediate temperatures have been shown, for example, in life history traits (Herrmann *et al.* in press) and metabolic rates (Chopelet *et al.* 2008) and might play a major role under natural conditions (Mitchell & Lampert 2000). A notable exception, the study by Williams *et al.* (2012), demonstrated that heat tolerant *D. pulex* clones were not only more heat resistant due to increased levels of hemoglobin expression (compared to heat-sensitive clones), but also that these increased of hemoglobin levels were maintained at non-stressful temperatures. This observation suggests adaptive canalized expression, i.e., loss of plasticity in expression levels, in a temperature-relevant gene. A comparable observation was made by Jansen *et al.* (2017), who analyzed subpopulations of *D. magna* across time (utilizing qPCR and a resurrection ecology approach) and reported loss of plasticity in expression levels under evolutionary constraints in a large proportion of temperature candidate genes.

While not assessing temperature-driven differences under intermediate conditions themselves, the above mentioned studies represent important resources that we used to assemble a set of temperature-relevant candidate genes. Specifically, temperature-relevant reference candidate genes were obtained from the literature on gene functions, thermal adaptation, and temperature-dependent expression rates in arthropods, associated with differences either in thermal tolerance or in thermal habitat properties. We compared transcriptomes among four populations of *Daphnia galeata* under non-stressful conditions utilizing RNA-seq, which allowed us to simultaneously estimate population divergence both on the sequence and the expression level. We applied several sequence- and expression level-based tests to find transcripts possibly under local selection or contributing substantially to population divergence. We expect an overrepresentation of temperature-relevant candidate genes among such transcripts if local thermal selection is a major factor in shaping divergence among these populations. Furthermore, we analyzed the contributions of expression level and coding sequence divergence to the proportion of population divergence that is potentially attributable to local thermal adaptation among populations.

Materials and Methods

Identification of temperature-relevant candidate genes and candidate gene transcripts

To obtain a set of reference candidate genes, we searched Web of Science (http://apps.webofknowledge.com) using different combinations of the keywords temperature, candidate genes, molecular, thermal selection, thermal adaptation, heat tolerance, phenotypic plasticity, freshwater zooplankton, expression, genome scan, heat stress, thermal genes, and RNA-seq. The results were screened for studies experimentally demonstrating temperature-relevance of genetic loci (reference candidate genes) and included literature on gene functions, thermal adaptation, and temperature-dependent expression rates in arthropods (Tables S1 and S2, Supporting Information C3). Amino acid sequences of 32903 *D. galeata* transcripts and 17014 reference candidate genes were assigned to existing orthogroups in the OrthoMCL database (http://orthomcl.org; currently containing genomes of 150 species) using the OrthoMCL tool (Li *et al.* 2003) with default settings. Transcripts orthologous to a reference candidate gene were deemed candidate gene transcripts (CGT).

Sampling and RNA collection, preparation and sequencing

Transcriptomes of twenty-four clonal *D. galeata* lines from four lakes (population J: Jordán reservoir in Czech Republic, populations M and LC: Müggelsee and Lake Constance in Germany, and population G: Greifensee in Switzerland, Table C3-1 and Figure C3-1C) were analyzed. The clonal lines were established using hatchlings from propagules collected from the sediment (see Henning-Lucass *et al.* (2016) for hatching conditions and Herrmann *et al.* (in press) for sampling conditions). Laboratory conditions prior to harvesting and RNA extraction are described in (Huylmans *et al.* 2016).

In total, 72 total RNA samples (six clonal lines from four lakes; three replicates per clonal line) were sent to the company GATC (Konstanz, Germany) for library preparation and sequencing. The total RNA was reverse transcribed in cDNA using random primers. Sequencing was carried out on an Illumina HiSeq 2000 (San Diego, USA), with 50 bp single-end reads. The pooling of up to eight libraries per lane was designed carefully to avoid block effects, and each library was sequenced on two to four different lanes.

Quality control and trimming

The sequencing output was pre-filtered with the Casava pipeline (Illumina, San Diego, USA). Subsequently, all reads with ambiguous bases were removed. The reads were then trimmed at the 3' and 5' ends. Bases with a quality score (phred) lower than 20 and reads shorter than 45 bp after trimming were discarded. The trimming steps were done using the Galaxy (Afgan *et al.* 2016) local instance of the Gene Center in Munich, Germany.

Mapping to reference transcriptome and read counts

Trimmed 50bp single-end reads were mapped to the *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). The required identity was increased from the default setting (65% identity) due to the high number of recent gene duplications that can be expected based on the *D. pulex* genome (Colbourne *et al.* 2011). For the same reason, all seeds were used to build the lookup table from the transcriptome and the highest sensitivity setting was used. Read counts were obtained from SAM files using a custom python script and discarding ambiguously mapped reads.

Differential gene expression

The R package DESeq2 (Love *et al.* 2014) was used to identify transcripts that were differentially expressed between populations. We first performed 6 pairwise comparisons: G vs. J, G vs. LC, G vs. M, J vs. LC, J vs. M, and LC vs. M. All *P*-values were adjusted for multiple testing using the Benjamini-Hochberg correction (Benjamini & Hochberg 1995) implemented in DESeq2. A transcript was considered significantly differentially expressed between a pair of populations when the contrast yielded an adjusted *P*-value (*padjust*) lower than 0.05 and a fold-change (FC) of more than 1.5 times.

For all differentially expressed transcripts, we computed the quartile coefficient of dispersion based on the mean read counts per clonal line after normalization. The dispersion coefficient is an estimate of transcript-wise expression variability across all samples. A higher value signifies a higher variability of expression. To compare transcript-wise expression variability between candidate gene transcripts and non-candidate transcripts, we conducted two-sample t-tests (using R) on the two groups of log-transformed dispersion coefficients.

SNP calling and filtering

Initial variant calls were made using GATK's HaplotypeCaller (DePristo *et al.* 2011) via local realignment of haplotypes. Samples were genotyped jointly using GATK's GenotypeGVCFs tool. A single vcf file was created using the GATK pipeline described in the section 'Variant Calling and Filtering' with the exception that the '-allSites' flag was switched on to include all non-variant loci for which there were data available. Besides true variation, these initial variant calls contain false positives due to systematic sequencing artefacts, mismapped reads, and misaligned indels. Such false-positive calls often (i) exhibit excessive depth of read coverage, (ii) show an allelic imbalance, (iii) occur preferentially on a single strand, (iv) appear in regions of poor read alignment and (v) arise in unusual close proximity to multiple other variants. Thus, the majority of such calls can be detected and rejected using filters based on the above observations (Laurent *et al.* 2016). Variants were removed with GATK'S VariantFiltration using the following set of criteria (with acronyms as defined by the GATK package): clusterWindowSize = 35. (ii) QD < 2.0. (iii) FS > 30.0.

Population structure and population-specific loci

We used principal components analysis (PCA) and discriminant analyses of principal components (DAPC, Jombart *et al.* 2010) to assess population structure. PCA and DAPC were conducted in R with the package ADEGENET 2.0.1 (Jombart & Ahmed 2011). Input data sets for both analyses were identical (Table C3-2). Only biallelic SNPs free of indels were included in the analyses while other filters were applied as described above. Six pairwise DAPCs were performed to obtain allele loadings and identify the SNPs contributing to divergence among each pair of populations, conserving > 99% of the variance in each data set.

To obtain SNPs that contributed highly to overall sequence divergence among populations, we applied an arbitrary 10 % cutoff to the allele loadings obtained in each pairwise comparison. The top 10% of SNPs contributing the most to population divergence in each pairwise comparison were analyzed further to identify highly divergent SNPs for each population, i.e., SNPs that contributed highly to population divergence in each of the three pairwise comparisons possible for each population (e.g., for G: "G vs. J", "G vs. LC", and "G vs. M"). The SNPs shared among all three comparisons for one population were labeled "highly divergent" for the respective population (Figure C3-2B). The lists of shared SNPs were obtained using the online tool jvenn (http://bioinfo.genotoul.fr/jvenn/). Transcripts that contained at least one highly divergent SNP were labeled "highly divergent transcripts".

Jvenn was used to obtain transcripts with distinctive expression levels in a particular population. The procedure is represented exemplarily for upregulated transcripts in population G (Figure C3-2A). A significantly differentially expressed transcript was labeled "distinctively expressed" (a) if it occurred uniquely among the four populations ("List 1", Figure C3-2A) or (b) if it occurred in each of the three pairwise contrasts for this specific population ("List 2", Figure C3-2A). List 1 was obtained by compiling a set of all significantly upregulated transcripts in any comparison with

another population for each population, and all four sets combined in a four-way Venn diagram. These "private" transcripts were significantly upregulated only in the respective population and exhibited equal expression levels or non-significant differences among the other populations. List 2 was obtained by compiling three sets of transcripts for each population. Each set contained all transcripts that were upregulated in this population compared to one of the other populations (e.g., for G: transcripts upregulated G in comparison to J, in comparison to LC, and in comparison to M). After combining the two lists, distinctive transcripts were used in further analyses. Distinctively expressed transcripts were identified separately for up- and down-regulated transcripts in each population.

Identification of loci under selection, gene functions, and abundances

We used McDonald-Kreitman tests (MKT, McDonald & Kreitman 1991) with Jukes-Cantor corrected diversity as implemented by Egea *et al.* (2008; http://mkt.uab.es/mkt) to test transcripts for selection among pairs of populations. To obtain alignments of transcript sequences, SNP calling data sets were filtered as described above but without the SNP-cluster filter and with a minor allele frequency cutoff of 5%. Beagle 4.1 (Browning & Browning 2007) was used to phase SNP calling data and a python script was used to parse the phased vcf-file to sample sequences in fasta format. Protein sequences of these transcripts were obtained using EMBOSS getorf (Rice *et al.* 2000) to detect and subsequently choosing the longest open reading frame. Protein sequences were aligned using MUSCLE (Edgar 2004) and transcript alignments were subsequently obtained using EMBOSS tranalign (Rice *et al.* 2000), by aligning corresponding nucleotide sequences based on the protein alignments. MKTs could be applied to 26 pairwise comparisons with alignments showing non-zero values in all four categories of nucleotide changes (nonsynonymous (D_n) and synonymous substitutions (D_s) between populations; and nonsynonymous (P_n) and synonymous polymorphisms (P_s) within populations, Tables C3-2 and C3-3).

Two F_{ST}-based outlier approaches, implemented in the programs BayeScan (Foll & Gaggiotti 2008) and LOSITAN (Antao *et al.* 2008), were used to identify loci under selection. For both outlier tests, SNPs were filtered for low quality ("QD < 2.0"), strand bias ("FS > 30.0"), minor allele frequency ("MAF < 10%"), SNP clusters, and indels. For LOSITAN analysis, SNPs with missing data in the alternative allele field (SNP calling) were also filtered. LOSITAN outlier testing was conducted using 500.000 simulations under the infinite allele model, confidence interval of 0.99, false discovery rate of 0.1, attempted F_{ST} of 0.182, subsample size of 12 (computed by LOSITAN), and simulated F_{ST} of 0.181. For tests with BayeScan, we used default settings and a prior odd of 100. A BayeScan test was conducted over all four populations and six pairwise tests were performed among populations. Bonferroni correction was used to account for multiple testing. Numbers of analyzed SNPs and corresponding transcripts in all analyses are listed in Table C3-2.

Gene ontology (GO) terms were obtained from UniProt (http://www.uniprot.org, The UniProt Consortium 2014) for all reference candidate genes that corresponded to outliers and highly

divergent or distinctively expressed CGTs. Abundance of orthogroups as well as of biological function and cellular components GO terms was assessed for DAPC, differential expression and LOSITAN results. Pairwise Mood's median tests were applied to test for significant differences in GO term and orthogroup abundance between the sets of highly divergent and distinctively expressed CGTs.

Overrepresentation analyses

Hypergeometric tests were used to test for overrepresentation of candidate gene transcripts in results of LOSITAN, DAPC, and differential expression analyses (R function phyper, R Core Team 2014). *P*-values obtained in overrepresentation analyses were Bonferroni-corrected for multiple testing. Transcripts occurring in the output of at least three different tests were also tested for overrepresentation. Transcripts possibly under selection according to BayeScan and MKTs were not included because of their low sample size.

Results

Identification of temperature-relevant candidate genes and candidate gene transcripts

We obtained 17014 reference candidate genes from 51 studies by screening literature on gene functions, thermal adaptation, and temperature-dependent expression rates in arthropods for studies which experimentally demonstrate temperature-relevance of genetic loci (Tables S1 and S2, Supporting Information C3). By assigning the reference candidate genes and *D. galeata* transcripts to existing orthogroups in the OrthoMCL database, we obtained 7707 different orthogroups, all of which contained *D. galeata* transcripts (Tables S3 and S4, Supporting Information C3). Of the 32903 distinct *D. galeata* transcripts analyzed in total, 47% were not assigned to any OrthoMCL orthogroup and 10% were assigned to orthogroups without any reference candidate gene. 14072 *D. galeata* transcripts (43% of all transcripts) were assigned to 5664 orthogroups containing both reference candidate genes and *D. galeata* transcripts and thus deemed CGTs.

Population structure

The PCA based on SNPs obtained from six clonal lines from four lakes showed distinct clustering of clones in accordance with their source populations (Tables C3-1 and C3-2, and Figure C3-1 A and C). Clusters of populations M, G, and LC were only separated along the second principal component (explaining 14.9% of the variance), whereas J was separated from the others along the first principal component (explaining 16.1% of the variance). Clones J3 and J2.1 were not as clearly grouped among other Jordán reservoir clones but instead placed midway between the clusters of populations J and LC.

The DAPC (Table C3-2) supported the clear genetic differentiation among the populations detected in the PCA (Figure C3-1B). The DAPC explained 100% of the variance with 23 principal component eigenvalues and three discriminant analysis functions. Results of the DAPC analysis

indicated that the optimal number of clusters of individual genotypes was four according to the Bayesian information criterion (data not shown), corresponding to the number of analyzed populations. Inferred group membership of individual genotypes determined from SNP data also corresponded to the genotypes' origin, with the exception of clonal line J2.1 from J, which was grouped into a cluster consisting of all clonal lines from LC ($P < 0.001$ for group membership of all genotypes).

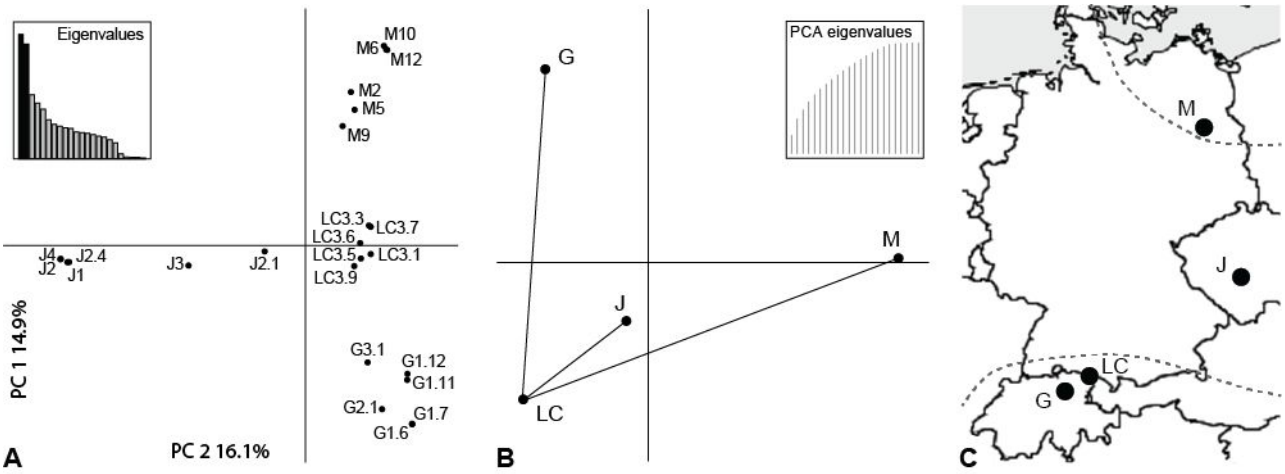

Figure C3-1: Genetic differentiation among *D. galeata* populations of Greifensee (G), Jordán reservoir (J), Lake Constance (LC), and Müggelsee (M). A: PCA and distribution of eigenvalues (filled black bars indicate eigenvalues corresponding to shown principal components); B: DAPC including a minimum spanning tree based on the squared distances between populations; C: distribution of sampling localities. Dotted grey lines mark the extent of the Alpine and British-Irish, and Scandinavian ice sheets during the last glacial maximum (Svendsen *et al.* 2004; Fitzsimmons & Hambach 2014)

Table C3-1: Sampling sites

Lake	Location	GPS coordinates	Alt. [m]	Vol. [km3]	Dep. [m]	Temp. [°C]
Greifensee	Switzerland	47° 21‘20“ N, 8° 40‘10“ E	435	0.1485	34	9.26
Jordán Reservoir	Czech Republic	49° 24‘55“ N, 14° 39‘49“ E	437	0.0027	14	12.56
Lake Constance	Austria, Germany, Switzerland	47° 37‘21“ N, 9° 26‘24“ E	395	48	254	10.32
Müggelsee	Germany	52° 26‘6“ N, 13° 38‘6“ E	32	0.0366	8	8.75

Alt.: altitude; Vol.; volume; Dep.: maximum depth; Temp.: average air temperatures near sampling sites in the years 2000-2010 according to the Climate Research Unit of the British Atmospheric Data Centre (Mitchell & Jones 2005; BADC 2008).

Table C3-2: Summary of transcript and SNP data sets used in various analyses; and detected outliers, distinctively expressed, and highly divergent transcripts

Analysis	# analyzed transcripts (SNPs)	# analyzed CGTs (SNPs)	# detected transcripts (SNPs)	# detected CGTs (SNPs)
BayeScan total	10783 (64287)	6857 (46966)	7 (7)	3 (3)
BayeScan all			5 (5)	1 (1)
BayeScan G-LC			2 (2)	2 (2)
MKT	25	15	2	2
LOSITAN	10760 (64237)	6854 (46954)	782 (997)	488 (640)
DAPC total	10138 (62858)	6656 (46306)	2064 (3566)	1433 (2832)
DAPC G			627 (982)	453 (790)
DAPC J			719 (1075)	507 (847)
DAPC LC			335 (470)	239 (380)
DAPC M			730 (1046)	496 (820)
DE	32903	14072	11450	4237
DE up			8264	2973
DE down			6597	2462
DE G up			2809	1007
DE J up			2114	758
DE LC up			1129	380
DE M up			2212	828
DE G down			1486	529
DE J down			2062	846
DE LC down			1296	464
DE M down			1753	623

MKT: McDonald-Kreitman test; DAPC: highly divergent transcripts according to discriminant analysis of principal components (the same input was analyzed in principal component analysis); DE: distinctively expressed transcripts; CGT: candidate gene transcript. # detected transcripts/CGTs: number of outliers, distinctively expressed or highly divergent transcripts or CGTs. Data sets marked in bold were tested for overrepresentation of CGTs. Both MKTs and analyses of differential expression were applied on the level of transcript sequences. In BayeScan, LOSITAN, and DAPCs, SNPs were analyzed that correspond to the indicated number of transcripts.

Expression-based population divergence – distinctively expressed transcripts

Over all populations, 11450 significantly differentially expressed transcripts were "distinctively expressed" (Table C3-2), i.e., up- or down-regulated only in one of the populations or exhibiting higher or lower expression levels in one population compared to all others (Figure C3-2A). We observed little difference in the proportions of distinctively expressed genes per population between the complete set of transcripts and the CGT subset (Table C3-2, Figure C3-3A). In the total set of distinctively expressed transcripts as well as in the set of upregulated transcripts, LC has by far the lowest number of transcripts (and CGTs), followed by M, G, and J, with little difference among values for M, G, and J (Figure C3-3A). This corresponds well to the population structure observed in the PCA and the DAPC minimum spanning tree (Figure C3-1 A and B), in which all

other populations were connected to LC. Thus, LC shows higher similarity to the other populations than they show to each other, both on the expression and on the sequence level. The order of transcript numbers differs in the downregulated subset of distinctively expressed transcripts, while also showing a more homogenous distribution. While LC remains the population with the lowest number of distinctively expressed transcripts, it is followed closely by G, M, and J. We also observed a more pronounced difference between the complete set of transcripts and the CGT subset. The difference in the amount of distinctively downregulated transcripts between population J and the other populations is substantially larger in the subset of CGTs that in the total data set (or in any of the other analyses; Figure C3-3).

Quartile coefficients of dispersion were computed as a measure of expression variability for 13717 CGTs and 18303 non-CGTs. For 883 transcripts, the dispersion coefficient could not be computed because of missing data in more than half of the 24 samples. The dispersion coefficients ranged from 142.93 (*Dgal_o1269t5*) to 0.03 in CGTs and from 226.26 (*Dgal_s329070*) to 0.04 in non-CGTs. The distribution of dispersion coefficient was massively skewed to the right in both data sets (77% of non-CGTs and 83% of CGTs exhibited values below one; and 99% exhibited values below ten in both data sets, data not shown). The means of the data sets, 0.81 for CGTs and 1.04 for non-CGTs, were significantly different (two-sample t-test, df = 32018, $P < 0.001$) but differ only by 0.23.

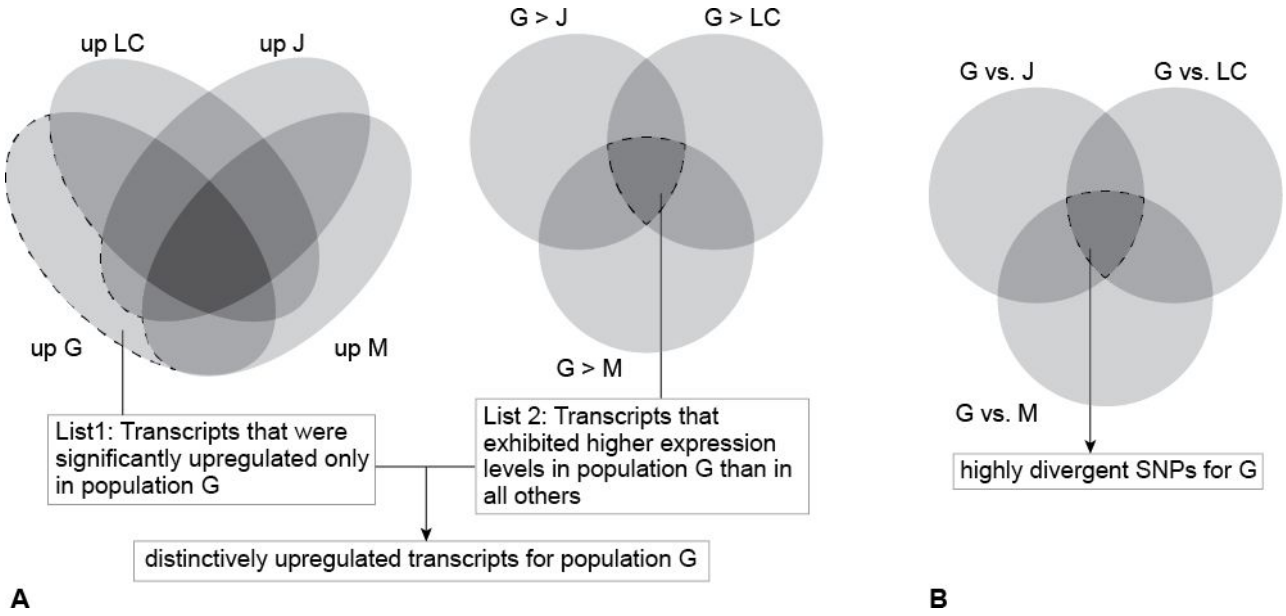

Figure C3-2: Identification of distinctively expressed transcripts (A) and highly divergent SNPs (B); exemplary shown for population G. A: Venn diagrams based on significantly differentially expressed transcripts (*padjust* < 0.05; FC > 1.5). The left Venn diagram contains the upregulated transcripts for each population (in at least on comparison with another population); the right Venn diagram contains transcripts upregulated in G compared to each other population. All transcripts found in at least one of the indicated intersections were labeled "distinctively expressed". For each population, distinctively up- as well as downregulated transcripts were identified. B: Venn diagram based on all SNPs exhibiting allele loadings within the top 10% contributing to cluster separation in pairwise DAPCs between population G and each other population (6285 SNPs per pairwise comparison). Transcripts containing at least one "highly divergent" SNP were considered "highly divergent transcripts"

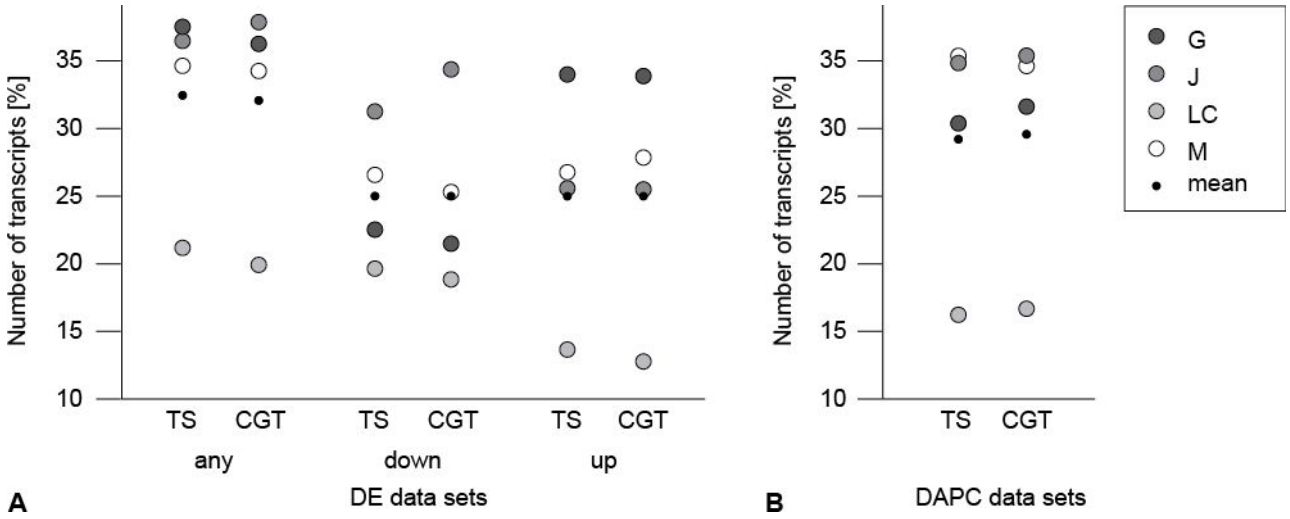

Figure C3-3: Population-wise proportions of distinctively expressed (A) and highly divergent (B) transcripts (relative to the total number of population-specific transcripts in each analysis). TS: all transcripts, CGT: candidate gene transcripts. A: any: all distinctively expressed transcripts, down: distinctively downregulated transcripts, up: distinctively upregulated transcripts. These transcripts were significantly differentially expressed (*padjust* < 0.05; FC > 1.5). B: highly divergent transcripts contain at least one SNPs exhibiting allele loadings within the top 10% contributing to cluster separation in pairwise DAPCs

Sequence-based population divergence – highly divergent loci

Based on the allele loadings of SNPs computed in six pairwise DAPCs between the four populations, we identified SNPs that contributed highly to overall sequence divergence specifically for each population in comparison with all other populations (Figure C3-2B). The highest amount of such "highly divergent" SNPs within CGTs was found in population J with 847 SNPs shared among all pairwise comparisons with other populations, corresponding to 507 highly divergent transcripts for population J (Table C3-2, Figure C3-3B). M and G followed with lower but similar amounts, 820 and 790 highly divergent SNPs respectively, corresponding to 496 and 453 highly divergent transcripts. With 380 highly divergent SNPs (corresponding to 353 highly divergent transcripts), LC has by far the lowest amount of highly divergent loci. This corresponds to the low amount of distinctively expressed transcripts in this population (Figure C3-3A) as well as to its central position in the PCA and the DAPC minimum spanning tree (Figures C3-2A and B).

We observed little difference between the complete sets of highly divergent transcripts for each population and the subsets of highly divergent CGTs (Figure 3B). The pattern of roughly similar amounts of highly divergent loci for G, J, and M, and substantially lower amounts for LC was also observed among highly divergent CGTs. It also corresponds well to the population-wise amounts of distinctively expressed CGTs (Figure 3A).

Loci under selection

MKTs could be applied to only 26 pairwise comparisons, in which transcript alignments showed non-zero values in all four categories of nucleotide changes (nonsynonymous and synonymous substitutions; and nonsynonymous and synonymous polymorphisms, Tables C3-2 and C3-3).

According to MKTs, sixteen transcripts possibly under positive selection in pairwise comparisons among populations (neutrality index NI < 1) were detected. Two of them, both candidate gene transcripts, exhibit significant *P*-values in spite of the low divergence among populations: *Dgal_t25857c0t3* in populations LC and M (NI = 0.157, $P < 0.01$) and *Dgal_t24940c1t3* in populations J and M (NI = 0.028, $P < 0.01$). The first is orthologous to an uncharacterized integral membrane component protein (GO identifier: GO:0016021, Table C3-5,) that has been shown to be differentially expressed in *Drosophila melanogaster* between 21 °C and 29 °C in tropical flies (from Panama) but not in flies from temperate climate (from Maine, Zhao *et al.* 2015). Also, decreased expression along a temperature gradient of 13, 18, 23, and 29 °C (Chen *et al.* 2015) and increased expression after heat stress (37 °C for 5.5 h, Landis *et al.* 2012) have been shown in *D. melanogaster*. The second is orthologous to a gene involved in axon guidance and protein localization (GO identifiers: GO:0007411, GO:0008104, GO:0005515) found to be differentially expressed among different temperatures in *D. melanogaster* in several studies (upregulated after a heat-shock of 37 °C for 5.5 h, Landis *et al.* 2012; DE among 20 and 29 °C, Vermeulen *et al.* 2014) . A study comparing the response of wild-type and cold-sensitive flies showed differential expression between 0 and 25 °C in both (0 and 25 °C, Vermeulen *et al.* 2013). Although both are CGTs and were differentially expressed in reference studies, they were not detected in any outlier test, highly divergent or distinctively expressed in any population.

Of the seven transcripts possibly under negative selection (NI > 1), none showed significant values. Only one transcript, *Dgal_o3239d43144t1*, was tested in two different pairwise comparisons (G vs. M and J vs. M). Three transcripts were neutral (NI = 1, $P > 0.99$). Interestingly, of 26 possible tests (on 25 transcripts), 23 involved population M and only two involved population LC. The reason for the low number of possible tests is that pairwise comparisons that reveal fixed substitutions in transcripts were extremely rare, with 280 found in 70362 pairwise comparisons (data not shown). A high amount of these fixed substitutions, 192, occurred between population M and one of the other populations, while fixed substitutions were much rarer among G, J, and LC. Fixed substitutions in population LC were only detected in 28 pairwise comparisons.

BayeScan analysis based on SNPs among all four populations (Table C3-2) indicated five SNPs as outliers based on F_{ST}-values in five different transcripts: *Dgal_o6007t2*, *Dgal_s329374*, *Dgal_sd36072361675*, *Dgal_t14377c0t1*, and *Dgal_s269867*. The latter is the only CGT among BayeScan outliers and also highly divergent for M. In pairwise BayeScan analyses of populations LC and G, two additional outlier SNPs in CGTs, *Dgal_t25421c0t1* and *Dgal_t25538c0t9*, were detected. These two transcripts were also highly divergent in pairwise DAPC analyses for LC and G, respectively. The other five pairwise tests revealed no outliers, although other population pairs exhibited a larger genetic distance, according to the PCA. All seven BayeScan outliers were either also outliers in LOSITAN (*Dgal_s329374*, *Dgal_sd36072361675*, *Dgal_t14377c0t1*), highly divergent

according to pairwise DAPC allele loading analyses (*Dgal_s269867*, *Dgal_t25421c0t1*) or both (*Dgal_o6007t2*, *Dgal_t25538c0t9*).

Table C3-3: Results of McDonald-Kreitman tests

Transcript ID	Pop1	Pop2	P_S	D_S	P_N	D_N	NI		CGT
positive selection									
Dgal_t24940c1t3	**J**	**M**	**14**	**1**	**2**	**5.01**	**0.03**	**0.97**	**yes**
Dgal_o3048t2	G	M	27	1	7	2	0.13	0.87	yes
Dgal_t25857c0t3	**LC**	**M**	**152**	**13.13**	**33**	**18.08**	**0.16**	**0.84**	**yes**
Dgal_o6811d29634t1	G	J	6	1	1	1	0.17	0.83	yes
Dgal_o1950d42751t1	J	M	31	1	8	1	0.26	0.74	
Dgal_t22867c0t1	G	M	38	1	10	1	0.26	0.74	
Dgal_t22541c0t4	J	M	15	1	5	1	0.33	0.67	yes
Dgal_o2921t5	J	M	5	1	2	1	0.40	0.60	
Dgal_t25680c0t1	J	M	72	1	34	1	0.47	0.53	yes
Dgal_o20756d33679t1	G	M	4	1	2	1	0.50	0.50	yes
Dgal_o2092t4	J	M	4	1	2	1	0.50	0.50	
Dgal_t19507c0t1	G	J	10	2	5	2	0.50	0.50	
Dgal_t26021c0t1	G	M	247	3	134	3	0.54	0.46	yes
Dgal_s383825	G	M	30	2.01	11	1	0.74	0.26	
Dgal_t25433c0t3	G	M	4	2.01	6	4.01	0.75	0.25	yes
Dgal_o3239d43144t1	G	M	13	1	12	1	0.92	0.08	yes
selectively neutral									
Dgal_ak44_j_645050	J	L	2	1	2	1	1.00	0.00	yes
Dgal_o3239d43144t1	J	M	13	1	13	1	1.00	0.00	yes
Dgal_s371227	J	M	4	1	4	1	1.00	0.00	yes
negative selection									
Dgal_o4720t4	J	M	1	5.1	2	9.1	1.12	-0.12	
Dgal_o6192d38232t1	J	M	25	4.05	17	2	1.38	-0.38	yes
Dgal_s180181	G	M	16	11.39	38	16.87	1.60	-0.60	
Dgal_o1089t4	G	M	4	2.02	7	2	1.76	-0.76	
Dgal_s388875	J	M	1	2	1	1	2.01	-1.01	yes
Dgal_a4274	J	M	1	1	3	1	3.01	-2.01	yes
Dgal_ak34_f_723347	G	M	3	3.04	4	1	4.05	-3.05	

Only transcripts with non-zero values in all four categories of substitutions were used. P_S: Number of silent polymorphic sites; D_S: silent divergence (Jukes-Cantor corrected); P_N: number of non-silent polymorphic sites; D_N: non-silent divergence (Jukes-Cantor corrected); NI: neutrality index; : proportion of adaptive substitutions (Smith & Eyre-Walker 2002); CGT: candidate gene transcript. Significant results ($P < 0.01$) are indicated in bold.

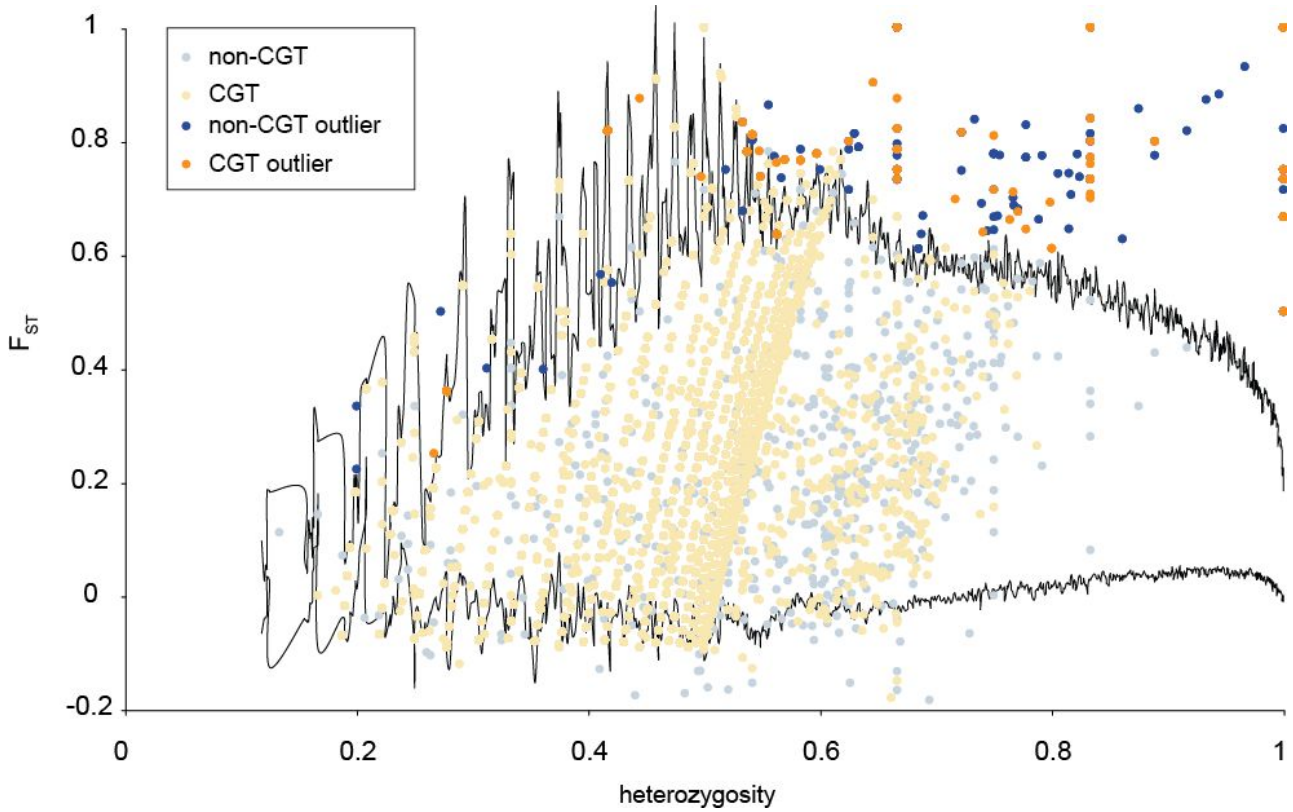

Figure C3-4: Outlier SNPs detected by LOSITAN. Lines indicate 99% and 1% confidence intervals. CGT: candidate gene transcript

Of the 64237 SNPs analyzed in LOSITAN (Table C3-2), 997 SNPs, corresponding to 782 transcripts, were detected as F_{ST}-based outliers under directional selection (Figure 4). Sixty-four percent of these (640 SNPs) corresponded to 488 candidate gene transcripts. Forty-one percent (199) of the LOSITAN CGTs outliers were either detected as outliers in BayeScan as well (only *Dgal_t25538c0t9*) or highly divergent based on allele loadings in pairwise DAPCs. We focused on this subset of "supported" LOSITAN outliers for further analyses because it is probably enriched for true positive outliers. An examination of the associated gene functions showed that supported LOSITAN outlier CGTs were orthologous to a wide variety of reference candidate genes associated with diverse GO terms. The vast majority of associated GO terms were also associated with distinctively expressed CGTs (84% of biological function GO terms and 81% of cellular components GO terms, Table C3-4).

In the category biological functions, the most abundant GO terms (Table S6, Supporting Information C3) were neurogenesis (GO:0022008), DNA-dependent transcription (GO:0006351), and protein phosphorylation (GO:0006468). In the category of cellular components, integral component of membrane (GO:0016021), nucleus (GO:0005634), and cytoplasm (GO:0005737) were most abundant (Table S7, Supporting Information C3).

Abundances of GO terms and orthogroups

The majority of gene functions was shared among subsets of highly divergent and distinctively expressed CGTs: 68% of the biological process GO terms associated to highly divergent CGTs were found among distinctively expressed CGTs, and 77% vice versa. Almost identical values were obtained for cellular component GO terms (68% and 76%, respectively; Table C3-4). The most frequent biological function GO terms differ among highly divergent and distinctively expressed CGTs (Table S6, Supporting Information C3). Highly abundant cellular component GO terms were, by contrast, almost consistent among both subsets (Table S7, Supporting Information C3). While many high-abundance GO terms of highly divergent CGTs were also highly abundant among distinctively expressed CGTs, the opposite is not true. For example the cellular component GO terms fusome (GO:0045169) and spectrosome (GO:0045170) were highly abundant in distinctively expressed CGTs but much less abundant in highly divergent CGTs.

Table C3-4: Summary of GO term associated outlier, highly divergent, distinctively expressed candidate gene transcripts

Analysis and GO category	# CGTs	# GO terms	# GO terms/TS	# CGTs shared with			# GO terms shared with		
				DE	DAPC	LOSI sup.	DE	DAPC	LOSI sup.
BP	3864	2972	-	-	-	-	-	-	-
DE	2851	2177	0.8	-	3.3%	0.6%	-	76.8%	34.3%
DAPC	1107	2468	2.2	8.5%	-	13.8%	67.8%	-	36.1%
LOSI sup.	153	890	5.8	11.1%	100%	-	83.8%	100%	-
CC	3946	597	-	-	-	-	-	-	-
DE	2926	441	0.2	-	3.5%	0.7%	-	76.0%	43.1%
DAPC	1122	491	0.4	9.1%	-	14.3%	68.2%	-	62.3%
LOSI sup.	160	190	1.2	13.1%	100%	-	80.5%	100%	-

GO: gene ontology; BP: biological process; CC: cell component; DE: distinctively expressed; DAPC: highly divergent transcripts; LOSI sup.: supported LOSITAN outliers (also detected by BayeScan or highly divergent); # CGTs: number of candidate gene transcripts (with associated GO terms); # GO terms/TS: number of GO terms per transcript.

The subset of highly divergent CGTs exhibits a higher diversity of biological function GO terms than the subset of distinctively expressed CGTs with mean numbers of 2.2 and 0.8 GO terms per transcript, respectively (Table S8, Supporting Information C3). Consequently, the mean abundance of GO terms among distinctively expressed CGTs was 2.6 times higher than among highly divergent CGTs. These differences in GO term abundances and diversity between the subsets were significant (pairwise Mood's median tests, $P < 0.001$; Table S9, Supporting Information C3). These results correspond to the analyses of orthogroup abundances. The mean number of different orthogroups per transcript was 0.86 for highly divergent and 0.27 for distinctively expressed CGTs, while the mean abundance of orthogroups was about three times higher in distinctively expressed CGTs than in highly divergent CGTs (Table S8, Supporting Information C3). Over all transcripts (not restricted to CGTs), this difference between the subsets remained (on average 0.74

orthogroups per transcript in highly divergent and 0.16 in distinctively expressed transcripts). Nonetheless, orthogroup abundances were significantly different between highly divergent and distinctively expressed subsets, both in all transcripts and in CGTs (pairwise Mood's median tests, $P < 0.001$; Table S9, Supporting Information C3).

Table C3-5: Hypergeometric tests for overrepresentation of candidate genes in test results

Analysis	CGT input [%]	CGT output [%]	difference [%]
LOSITAN	63.7	62.4	1.3
DAPC G	**65.7**	**72.3**	**6.6****
DAPC J		**70.5**	**4.9***
DAPC LC		71.3	5.7
DAPC M		68.0	2.3
DE up	42.8	36.0	6.8
DE down		37.3	5.4
DE G up		35.8	6.9
DE G down		35.6	7.2
DE J up		35.9	6.9
DE J down		41.0	1.7
DE LC up		33.7	9.1
DE LC down		35.8	7.0
DE M up		37.4	5.3
DE M down		35.5	7.2

DAPC: highly divergent transcripts; DE: distinctively expressed transcripts; CGT: candidate gene transcripts. Significant results are marked in bold.

*$P < 0.05$; **$P < 0.01$

Overrepresentation analyses

Hypergeometric tests were used to test for overrepresentation of candidate gene transcripts in results of LOSITAN, DAPC, and differential expression analyses and revealed that the proportion of CGTs among LOSITAN outliers and distinctively expressed transcripts was lower than among the input data sets (Table C3-5). In highly divergent transcripts according to pairwise DAPCs, significant overrepresentation of CGTs was found for populations G (6.6% higher than expected, hypergeometric test, $P < 0.01$) and J (4.9% higher than expected, $P < 0.05$).

Discussion

In this study, we assessed temperature-related population divergence among natural populations of *D. galeata* at the expression as well as at the DNA sequence level. We aimed at identifying possible target genes for adaptation to different thermal regimes and to assess the role of temperature in

shaping divergence among populations. Therefore, we focused on transcripts orthologous to candidate genes for thermal adaptation collected form a broad variety of studies on arthropods.

Utilizing sets of stressor-specific genes in combination with genome- or transcriptome-wide markers represents a particularly useful approach. It provides the opportunity to assess the impact of important environmental factors on population divergence and allows for testing of specific hypothesis across a wide range of factors as well as organisms, including non-model organisms. A general difficulty, however, particularly in non-model organisms, is that species-specific transcripts are not assessed when they are not annotated (either function, importance or both are unknown). In the first *D. pulex* genome (Colbourne *et al.* 2011), more than 36% of the protein coding genes were not assigned to orthologous genes in insects (Huylmans *et al.* 2016). These genes of unknown function were suggested to play an important role in local adaptation and phenotypic variation (Colbourne *et al.* 2011; Jansen *et al.* 2017). However, the number of genes of unknown function in *D. pulex* might be inflated and is probably not representative of the *Daphnia* genus (Ye *et al.* 2017). In the *D. pulex* genome assembly recently published by Ye *et al.* (2017), only 28% of annotated genes were without a homolog in *Caenorhabditis elegans*, *Strigamia maritime*, *Anopheles gambiae*, *D. melanogaster* or *Homo sapiens*, compared to 57% of *D. pulex* genes without a homolog in these species in the genome published by Colbourne *et al.* (2011). Nonetheless, species-specific genes will inevitably be missed to a certain degree, even when, like in the present study, a large number of reference studies is available. In the present study on *D. galeata*, 47% of analyzed transcripts were not assigned to an OrthoMCL orthogroup and their functions need to be assessed separately.

Based on the 43% of transcripts successfully assigned to an orthogroup together with a reference candidate gene, we found that temperature candidate gene transcripts were overrepresented in two discriminant analyses of principal components; i.e., among transcripts that were highly divergent in populations of Greifensee and Jordán reservoir (based on DNA sequences). Therefore, temperature likely contributes to coding sequence divergence among the sampled populations, a finding also supported by McDonald-Kreitman tests. Although MKTs were not included in overrepresentation analyses because of the low sample size, both significant MKT results were CGTs. Temperature-related candidate genes were, by contrast, not overrepresented among distinctively expressed transcripts, although several studies suggest that phenotypic plasticity plays an important role in thermal adaptation in *Daphnia* (reviewed in Stoks *et al.* 2014). Among populations experiencing different thermal conditions, phenotypic plasticity is likely under divergent thermal selection, since it often results (like other phenotypic variation) from heritable variation in expression levels (Wray *et al.* 2003; Romero *et al.* 2012; Pai *et al.* 2015). However, the lack of CGTs overrepresentation at the expression level suggests that temperature is, at least under stress-free common garden conditions, not a major contributor to divergence in gene expression levels among populations. This is supported by the results obtained from our analysis of expression variability. In genes under evolutionary constraints in *Daphnia*, canalization of gene expression was observed repeatedly (Yampolsky *et al.* 2014b; Jansen *et al.* 2017). Although we found a

significantly lower mean variability in CGTs compared to the complete set of all transcripts, the actual difference between the means (0.23) was marginal and is thus not indicative of canalized gene expression in temperature-relevant candidate genes. While the majority of transcripts (nearly 80%) exhibited very low dispersion coefficients (lesser than one), transcripts with highly variable expression were found in both the candidate gene and the non-candidate gene subset. This also suggests that the evolutionary potential of the analyzed *D. galeata* populations regarding temperature is not depleted so far.

Analyses of population structure and the population-wise amounts of distinctively expressed and highly divergent transcripts revealed that a phylogeographic signal might conceal a weaker signal of selection. In the population from Lake Constance, we found by far the lowest amount of distinctively expressed and highly divergent transcripts and therefore the lowest divergence among the studied populations, both at the expression and at the sequence level. The low number of distinctively expressed genes in population LC signifies an excess of averagely expressed genes. On one hand, this might be explained by selectively neutral processes: the populations from Müggelsee, Greifensee, and Jordán reservoir might share LC as their common source population. Since Lake Constance is located at a lower altitude than Lake Greifensee and further south compared to Müggelsee and Jordán Reservoir, it was probably the first ice-free lake shortly after the last glacial maximum and thus very likely harbors the oldest of the studied populations (Figure 1C). The Jordán reservoir is an artificial reservoir only build in the 15th century (Kubecka & Bohm 1991) and was thus colonized last. A phylogeographic signal explains the lower amount of DNA sequence and expression divergence observed in LC and is also in accordance with a recent study on *D. magna* that demonstrated little change in population structure even following severe artificial selection (Jansen *et al.* 2017). Therefore, weak signals of selection are possibly concealed.

On the other hand, an ecological explanation is just as likely. The high amount of averagely expressed genes is possibly indicative of a homogenous and intermediate environment in Lake Constance compared to the other habitats. Since newly selected alleles are more likely to be fixed in heterogeneous environments, when either migration is low (Gavrilets & Gibson 2002; Vuilleumier *et al.* 2010) or follows an island model among populations of relatively homogeneous sizes (Joost *et al.* 2013), the low number of highly divergent CGTs might also be indicative of an homogenous environment. Lake Constance is by far the largest of the studied lakes (Table C3-1) and has the highest capacity of buffering against environmental change in general and in particular against thermal fluctuation (Carvalho 1987).

Although overrepresentation analyses suggest that temperature CGTs are not the main source of population divergence, many of them showed signals of selection, distinctive expression levels or high levels of DNA sequence divergence. These transcripts may contribute to local adaptation to temperature. Therefore, the elevated amount of distinctively downregulated transcripts observed in J (compared to the other populations) might signify a signal of selection and possibly a temperature

effect. Jordán reservoir is the habitat with the highest mean annual air temperature over the last decade and, because of its small water body (0.006% - 7% of the volume of the other lakes), probably experiences the largest temperature fluctuations throughout the year (Table C3-1). However, further evidence is needed to confirm this hypothesis.

An ecological explanation seems also reasonable for the high amount of MKTs that involved comparisons with population M (23 of 26 possible tests), or, more precisely, for the high number of transcripts that exhibited fixed substitutions in population M in comparisons with the other populations. The overall low level of population structure (among the four populations) at the transcriptome level and the consequently very low amount of transcripts with fixed substitutions (228, data not shown) resulted in the low number of performed MKTs (N = 26; 0.04% of theoretically possible tests), 16 thereof involving CGTs. The MKT is a conservative test for selection (Hudson *et al.* 1987) and low divergence, together with the lack of introns (i.e., a reduced amount of neutral diversity) also resulted in low power of the performed tests. The strikingly high amount of transcripts with fixed substitutions observed in population M likely reflects that Müggelsee is ecologically distinct from the other lakes. It is extremely shallow compared to the other lakes, the most northern of the lakes and located at the lowest altitude (Table C3-1). LC, by contrast, exhibited only an extremely low amount of transcripts with fixed substitutions, which is in accordance with the proposal of LC as possible origin of the other populations (Fig 1 B).

The two significant MKT results, positive selection at *Dgal_t24940c1t3* and *Dgal_t25857c0t3*, were both CGTs. Both were shown to be differentially expressed in a variety of reference studies in response to a broad spectrum of thermal conditions (Landis *et al.* 2012; Vermeulen *et al.* 2013; Vermeulen *et al.* 2014; Chen *et al.* 2015; Zhao *et al.* 2015). Therefore, these genes are most likely part of the general cellular stress response, rather than involved in specifically temperature-related responses. In contrast to observations in several reference studies, both transcripts were not differentially expressed under non-stressful conditions among the populations in the present study.

Selection acting on all three CGTs among BayeScan outliers (*Dgal_s269867*, *Dgal_t25421c0t1*, and *Dgal_t25538c0t9*) was supported by pairwise DAPC allele loadings, whereas only the latter was also an outlier LOSITAN. *Dgal_t25421c0t1* and *Dgal_t25538c0t9* were outliers in pairwise BayeScan tests between populations LC and G, and they were highly divergent in DAPC analyses for LC and G, respectively. These two populations also showed divergent development times (time to first reproduction, TTR) at non-stressful common garden conditions (TTR: 15 °C, LC > G) in life-history experiments (Herrmann *et al.* in press). Furthermore, the clonal lines from Greifensee and Lake Constance analyzed in the present study exhibited significantly different somatic growth rates (SGR: 20 °C, G > LC) and LC clones showed a trend for a higher reproduction rate according to preliminary results of life-history experiments (Verena Tams, Personal Communication). The gene orthologous to *Dgal_t25421c0t1* is involved in cell proliferation and was shown to be differentially expressed among inbred *D. melanogaster* lines with different levels of temperature sensitivity (heat-

sensitive, cold-sensitive, and control) in response to different thermal regimes (FlyBase ID: FBGN0015376; Vermeulen *et al.* 2013; Vermeulen *et al.* 2014; Table S5, Supporting Information C3). As such, this gene might contribute to the differences in life-history traits observed among G and LC under non-stressful common garden conditions (Herrmann *et al.* in press).

Expression level variation has long been recognized as an important component of the genetic basis for adaptation in addition to DNA sequence variation (Wray 2007). In studies on strains of the dinoflagellate *Alexandrium minutum* (Le Gac *et al.* 2016) and the pine species *Pinus contorta* and *Picea glauca x engelmannii* (Hodgins *et al.* 2016), for example, consistent patterns of sequence and expression divergence were found among the studied groups. If negative pleiotropy impairs adaptation on one of these levels, the other might constitute an alternative route for responding to selection pressures (Pai *et al.* 2015). Of the 32903 transcripts analyzed in the present study, 12.9% were CGTs and distinctively expressed in one of the populations. The number highly divergent CGTs in pairwise DAPC analysis (contributing substantially to population divergence in all three comparisons) was only a third of this amount (4.4%). Although we have no information on the phenotypic effects of variation on both levels and thus cannot precisely evaluate their relative contributions to individual fitness and local adaptation, these above mentioned observations indicate a stronger influence of thermal selection on the expression level. Simultaneously, only a small percentage of CGTs are shared among both gene sets (3.2% of distinctively expressed CGTs are highly divergent and 9.5% vice versa, Table S8, Supporting Information C3), which demonstrates the importance and interplay of both sequence and expression variation in shaping population divergence.

The interplay of responses to selective pressure on both the level of DNA sequence and gene expression often results in a direct association of genetic variation with variation in expression levels (Pai *et al.* 2015). Neither the transcripts under selection according to MKTs, nor the BayeScan outlier GCTs were significantly differentially expressed in the present study. The presence of coding sequence variation in absence of expression level variation indicates that local selection acts on the sequence level at these genes but does not affect their constitutive expression. Indeed, point mutations in protein coding sequences have been shown to result in differences of thermal optima among gene products (Arnold *et al.* 2001). It might, however, also reflect local selection acting on the expression level under specific conditions. Although we probably missed a major share of *trans*-regulatory sequences by applying a transcriptomic approach, expressed genes likely contain at least one *cis*-acting regulatory element (Pai *et al.* 2015) and regulatory elements can also be found within coding regions (e.g., Khan *et al.* 2012). Therefore, sequence divergence at these loci might signify selection acting on either the protein function or properties or, via *cis*-regulatory effects, on (facultative) gene expression. In genes showing adaptive differences in expression response to temperature stress, examples for both concurrent significant sequence differentiation (e.g., in *Cyp6a17* and *CG1304*, Turner, *et al.* 2008; Levine, *et al.* 2011) as well as a lack of sequence variation (e.g., in heat shock proteins, Schoville *et al.* 2012) have been found. The

15 CGTs associated with "heme-" or "hemoglobin"-related gene ontology terms in the present study (Table S10, Supporting Information C3) might represent an example for the latter case, i.e., for local thermal selection acting on to the expression levels but less on the coding sequence level. Interestingly, only four such GO terms were associated with any CGT (hemoglobin import, heme biosynthetic process, heme oxidation, and heme metabolic process), although the set of reference candidate genes contained more than 200 genes associated with 11 GO terms containing these keywords (data not shown). Heme-related CGTs were either distinctively expressed or highly divergent for one of the populations based on genetic sequences (according to either LOSITAN or pairwise DAPC analysis), but never both. Hemoglobin plays a major role in temperature sensitivity in *Daphnia* (e.g., Seidl *et al.* 2005; Zeis *et al.* 2013) and was suspected to contribute to local thermal adaptation (Yampolsky *et al.* 2014a). Since the involved loci were outliers in one test at most, with no further support, an impact of coding sequence variation is probably low. However, the distinctively expressed heme-related CGTs may indicate local thermal adaptation in the populations studied here.

Differences observed in GO term frequencies among highly divergent and distinctively expressed CGTs signify a general difference between CGTs that contribute highly to sequence-based population divergence on one hand and CGTs that contribute highly to expression-based population divergence on the other hand. Highly divergent CGTs were associated to a significantly wider variety of different GO terms with a lower frequency compared to distinctively expressed CGTs (for both cellular components and biological process GO terms, Mood's median tests, $P < 0.001$). The same phenomenon was observed for orthogroup abundances. This is interesting with regards to the differences we found between the total amounts of population-specific CGTs discussed above, with a four times higher amount of distinctively expressed CGTs compared to highly divergent CGTs. It indicates that distinctive expression profiles of temperature candidate genes (under stress-free conditions) are limited to a specific and reduced set of biological functions, while sequence divergence is to a lesser extent limited to this set of biological functions. This contrast may be a consequence of varying constrictions that affect divergence on the sequence and on the expression level. Sequence divergence is, for example, influenced by genomic architecture that can account for local reductions of effective population size. Inversions and genomic regions around centromeres or genes under selection show such local reductions of effective population size due to elevated rates of genetic drift in consequence of their typically low recombination rates (Pereira *et al.* 2016). Constrictions on the level of expression level include, for example, pleiotropic effects: changes in regulatory elements often affect the expression levels of several genes, but may also affect hundreds to thousands of genes (Paaby & Rockman 2013). Since our information is limited to expression levels in stress-free conditions, further studies are required to reveal the true contributions of sequence and expression based population divergence. Alternatively, the higher proportion of highly abundant GO terms and orthologous groups may signify an excess of duplicated genes in the subset of distinctively expressed CGTs. However, obtaining the full

D. galeata genome is necessary to verify this hypothesis. In addition, it would be beneficial to analyze transcriptomic responses across different temperature regimes to fully assess the relative contribution of thermal selection on the regulatory and on the coding level to understand the extent to which sequence or expression divergence is associated with different sets of gene functions.

While selective agents seem to target different sets of temperature candidate genes on the sequence and on the expression level, they target largely the same functions, however, to different degrees (Table C3-4; Tables S6, S7, and S8, Supporting Information C3). Notably, only 13% of CGTs under divergent selection according to outlier tests and simultaneously highly divergent are also distinctively expressed, while over 80% of GO terms associated with these CGTs were shared among both data sets. The most frequent cellular components GO terms were similar in all three data sets despite great discrepancies in absolute abundances. Besides nucleus, cytoplasm, and cytosol, they included several membrane-related terms. This observation reflects the important role of biological membranes in body temperature modulation. The fatty acid composition affects the amount of ion leakage across membranes as well as the activity of membrane-bound proteins. Therefore, membrane restructuring is a key mechanism in coping with climate variability and a major factor influencing the response time needed to accommodate temperature fluctuations (reviewed in Seebacher *et al.* 2010). Frequencies of biological process GO terms, however, were much more diverse among the data sets and relative frequencies were much lower in highly divergent CGTs (according to pairwise DAPCs) and LOSITAN outlier CGTs than in distinctively expressed CGTs (Table S6 and S8, Supporting Information C3). There was only one biological process GO term, protein phosphorylation, with a consistently high abundance and similar ranking in all three data sets, highlighting the importance of post-transcriptional regulation in population divergence. Protein phosphorylation is a post-translational modification that alters the structural conformation of proteins and thus causes activation, deactivation or alteration of their functions. It is suspected to play an important role for example in rapid cold hardening in *Drosophila* (Teets & Denlinger 2016) and in cellular responses to heat and cold stress in *Saccharomyces cerevisiae* (Kanshin *et al.* 2015). The high amount of CGTs related to protein phosphorylation and transcription in LOSITAN outliers thus may be indicative for adaptive plasticity in response to divergent temperature regimes among the populations, both on the transcriptional and on the post-transcriptional level.

In summary, our utilization of ecologically informative candidate genes is a useful approach with high flexibility and a wide range of potential applications. It revealed that temperature is probably not a major selective agent among *Daphnia* populations on the local scale, since we found no overrepresentation of CGTs in distinctively expressed transcripts and transcripts indicated as directionally selected. The population structure observed among the four populations can be explained by either local adaptation, a phylogenetic signal, which indicates LC as the source population, or a combination of both. Temperature seems, however, to contribute to population

divergence on the coding sequence level, as supported by the overrepresentation of highly divergent CGTs in the populations G and J. The high amount of fixed substitutions detected in population M, which is ecologically distinguished from the other populations due to its shallow depth and geographical location, also indicates an influence of temperature or other environmental factors. While the set of genes contributing highly to population divergence at the coding sequence level differs from the set of genes contributing to population divergence at the expression level, we observed that the associated gene functions were largely consistent. This supports the hypothesis that both constitute an alternative route for responding to selection pressures (Pai *et al.* 2015). Furthermore, a large amount of temperature-relevant candidate genes was distinctively expressed, outliers or highly divergent for one of the populations. These genes may contribute to local thermal adaptation and, together with the associated gene functions, constitute a valuable starting point for future research.

Acknowledgements

The authors would like to thank I. Schrank, L. Theodosiou, M. Kredler, C. Laforsch, J. Wolinska, J. Griebel, R. Jaenichen, and K. Otte for providing the necessary resources and help for maintaining *Daphnia* cultures in the laboratory. H. Lainer supported us for the molecular laboratory work. We thank C. Özsayar for providing invaluable bioinformatics and computing support. This study was supported by individual grants from the Volkswagen Stiftung (to MC; grant number 86030). This work benefits from and contributes to the Daphnia Genomics Consortium.

Data Accessibility

Raw sequence reads used for this study will be deposited in the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena).

The experimental set up for the analysis of differentially expressed genes will be available in ArrayExpress (https://www.ebi.ac.uk/arrayexpress).

Supporting Information

This information is available from the attached CD-ROM or from the online version of the article upon publication.

Table S1 List of references providing temperature-relevant reference candidate genes in arthropods.

Table S2 Reference candidate gene IDs or gene product IDs.

Table S3 Overview OrthoMCL orthogroup assignment.

Table S4 List of CGT orthogroups and reference candidate genes.

Table S5 List of GO terms corresponding to outliers, highly divergent, and distinctively expressed CGTs.

Table S6 Gene ontology terms (biological process) associated to distinctively expressed candidate gene transcripts (DE).

Table S7 Gene ontology terms (cellular component) associated to distinctively expressed candidate gene transcripts (DE).

Table S8 Overview of transcripts and orthogroups shared among data sets.

Table S9 Results of Mood's median tests.

Table S10 "heme-" or "hemoglobin" associated CGTs.

Author Contributions

MH and MC planned the study and designed the analyses; MC conducted the molecular work; MH, MC, and SPR analyzed the data; MH wrote the manuscript; all authors commented on results and contributed substantially to revisions of the manuscript.

General discussion

In my thesis, I focused on the traces of thermal selection in the genomes of *Daphnia* populations. Therefore, I employed a candidate gene approach, analyzing populations of the *D. longispina* species complex, comprising *D. galeata*, *D. longispina*, and their interspecific hybrids, across space and time (**Chapter 1**). I contrasted presumably neutral genetic population differentiation with differentiation at three putatively temperature-relevant candidate genes, *ERNA*, *MHC-1*, and *TRY5F*, and found signs of selection and possible thermal adaptation at candidate gene *TRY5F*. To further elucidate these results, I assessed the genotype-phenotype relationship of the three candidate genes in *D. galeata* populations (**Chapter 2**). I observed a general temperature effect and inter-population differences in phenotypic traits in life-history data. Furthermore, I revealed genotype-environment (GxE) interactions for several life-history traits at all candidate loci, which implies a possible contribution of candidate genes to several, potentially multilocus-based life-history traits. Lastly, I concentrated on the impact of local thermal selection on transcriptome-wide population divergence in *D. galeata*. I utilized a comprehensive set of potentially temperature-relevant candidate genes in arthropods and considered both expression level and coding sequence differences (**Chapter 3**). The results suggested that local temperature regimes were not a major force in shaping genetic and gene expression divergence among populations. Furthermore, no substantial reduction of expression level diversity was detected in the set of candidate genes compared to transcriptome-wide diversity. However, I found many temperature-related candidate genes with distinctive, population-specific expression levels or patterns of sequence divergence. Thus, I conclude that genetic variation was not specifically depleted in temperature-related gene expression in the sampled populations. These results bear important implications for my initial questions 1) Are natural populations of the of the *D. longispina* species complex adapted to local thermal regimes?, 2) Are interspecific hybridization and introgression contributing to local thermal adaptation?, and 3) What conclusions can be drawn regarding the fate of the *D. longispina* complex in light of advancing climate change?

Local thermal adaptation over space and time

Local adaptation has been hypothesized to play an important role in shaping population divergence in aquatic species that exhibit high dispersal capabilities, rapid population growth and the ability to produce resting propagules, like *Daphnia* species (De Meester *et al.* 2016). According to the Monopolization Hypothesis (De Meester *et al.* 2002), local adaptation in combination with founder and priority effects prevents gene flow and promotes the high levels of genetic differentiation, which are often observed among *Daphnia* populations (e.g., Hebert *et al.* 1993; Vanoverbeke & De Meester 1997; Palsson 2000; Ishida & Taylor 2007a; Thielsch *et al.* 2009). Adaptive change in

Daphnia in response to thermal selection was shown by several studies; in selection experiments (Van Doorslaer *et al.* 2009a; Van Doorslaer *et al.* 2010) as well as over time in natural populations (Geerts *et al.* 2015; Henning-Lucass *et al.* 2016). The results I presented in **Chapter 1** and **Chapter 3** support the occurrence of local adaptation among the analyzed populations and indicate potentially underlying genetic loci. The signs of local selection at candidate gene *TRY5F*, the correlation between genetic differentiation at this locus and temperature differences among populations as well as the lack of functional genetic divergence between *D. galeata* and *D. longispina*, all inferred from the analyses described in **Chapter 1**, might reflect local thermal adaptation. The many transcripts with distinctive expression profiles in each population and under selection according to outlier tests (**Chapter 3**) also indicate local adaptation, including local thermal adaptation. The latter was signified by the large proportion of temperature-related candidate genes among outlier or population-specific transcripts.

The signal of genetic differentiation I observed over time was, however, weak and not significant (**Chapter 1**). Additionally, a direct association among *TRY5F* alleles and phenotypic traits under different thermal conditions was not inferred, although several GxE interactions implied a contribution of each of the candidate genes to phenotypes under different thermal regimes (**Chapter 2**). Moreover, no genotype-phenotype relationships are known for *Daphnia* regarding the candidate genes possibly contributing to local thermal adaptation I described in **Chapter 3**. However, the observation of change over time as well as the observation of a fitness advantage through a putatively adaptive allele are important prerequisites for evidencing adaptive genetic change (Hansen *et al.* 2012). Therefore, further studies over time and on genotype-phenotype relationships are needed to elucidate the role of the candidate loci in local thermal adaptation.

The importance of thermal regulation and consequently of thermal adaptation in the face of climate change in organisms in general and specifically in aquatic ectotherms is stressed in a number of publications (e.g., Gibert & De Jong 2001; Hoffmann & Sgro 2011; Gunderson & Stillman 2015). The high proportion of temperature candidate genes I detected in the transcriptome of *D. galeata* (**Chapter 3**) - more than 40% of all analyzed transcripts and more than 60% of SNP-containing transcripts were candidate gene transcripts - reflects the importance of thermal regulation. Contrastingly, overrepresentation analyses were not suggesting a major influence of temperature in shaping population divergence. This large amount of temperature related candidate genes in the transcriptome without overrepresentation among genes possibly important in local adaptation might reflect an inflated number of candidate gene transcripts in *D. galeata* due to the orthologous grouping utilized to identify candidate transcripts. That is, gene functions of *D. galeata* transcripts might differ from their homologues identified in other arthropod species based on sequence similarity. It might, however, also reflect a strong influence of global rather than of local selection. Global selection would result in expression of a large amount of temperature-relevant genes but not necessarily in substantially elevated population divergence at such genes (compared to the total transcriptome) over the geographic scale included in my

analyses. Comparable studies on other populations, species, and selective factors are needed to assess whether the proportions of candidate genes I described in **Chapter 3** are to be expected or unusual.

Interspecific hybridization and introgression

Hybridization and introgression have been hypothesized to facilitate adaptation for quite some time and are by now recognized as important creative forces in plants as well as in animals (Arnold 1997; Rieseberg *et al.* 2003), particularly in disturbed environments (Chunco 2014). Hypothesis on potential consequences of environmental disturbance for *Daphnia* are, however, still controversial. While De Meester *et al.* (2016) suggest that anthropogenic disturbance might promote monopolization by creating spatial isolation and elevating the time lag between the arrival of immigrant genotypes, an opposite effect is also possible: Since environmental disturbance can eradicate the advantage of locally adapted genotypes and thus weaken an established monopolization of resources, it might promote effective gene flow and an influx of alien genotypes (Thielsch 2012). Subsequent differential introgression at functional loci may contribute to rapid local adaptation processes in altered environmental conditions and to the persistent permeability of reproduction barriers in *Daphnia* species complexes. My results, presented in **Chapter 1**, i.e., a lack of functional, but not of neutral genetic divergence between *D. galeata* and *D. longispina*, strongly support this latter hypothesis. The observed patterns of population differentiation are most likely explained by introgression at functional, protein coding, and putatively ecologically relevant candidate genes, including the gene *TRY5F*, which might contribute to local thermal adaptation. Prior to this study, examples of introgression in *Daphnia* were exclusively inferred from neutral genetic markers, including mitochondrial DNA (e.g., Cristescu *et al.* 2012; Markova *et al.* 2013), nuclear loci coding for ribosomal RNA internal transcribed spacers (ITS; e.g., Ishida *et al.* 2011), and microsatellite DNA (e.g., Brede *et al.* 2009; Alric *et al.* 2016).

Further work is required to assess the extent and direction of introgression among species of the *D. longispina* complex as well as the genetic regions involved, since incomplete lineage sorting could not be completely excluded as a source for the observed discrepancy between functional and neutral genetic divergence between *D. galeata* and *D. longispina*. However, based on the extensive literature on interspecific hybridization in the *D. longispina* complex (Weider & Wolf 1991; Schwenk 1993; Boersma & Vijverberg 1994; Schwenk & Spaak 1995; Jankowski & Straile 2003; Keller & Spaak 2004; Seidendorf *et al.* 2007; Yin *et al.* 2010; Griebel *et al.* 2015; Griebel *et al.* 2016) and numerous examples of introgression events (Taylor & Hebert 1993; Jankowski & Straile 2004; Taylor *et al.* 2005; Brede *et al.* 2009; Ishida *et al.* 2011; Rellstab *et al.* 2011; Alric *et al.* 2016), I consider adaptive introgression the most likely explanation.

The fate of the *D. longispina* species complex in light of advancing climate change

As aquatic ectotherms and inhabitants of freshwater ecosystems, *Daphnia* populations are particularly threatened by climate change (Gibert & De Jong 2001; Gunderson & Stillman 2015). Under ongoing selective pressure, their genetic variation might become depleted, which would severely limit the adaptive potential within populations and render them unable for further adaptive change (e.g., Hoffmann & Sgro 2011; Kelly *et al.* 2012a). As described in **Chapter 2**, I observed, indeed, significantly depleted genetic variation within experimental clones compared to samples from the wild at candidate gene *TRY5F*. The former clones were kept in the laboratory for several months to years prior to genotyping, while the latter were analyzed directly after isolation from lake sediment. The sediment clones thus represent a random sample of the natural diversity within these populations. Although *TRY5F* is differentially expressed under different temperatures in *D. pulex* (Schwerin *et al.* 2009) and *D. galeata* (M. Cordellier, Personal Communication), and probably under local thermal selection according to my analyses described in **Chapter 1**, the loss of genetic variation at this locus was not reflected in the phenotypic variation observed within and among *D. galeata* populations (**Chapter 2**): Strong phenotypic plasticity was shown, e.g., in somatic growth rate and time to first reproduction in response to different temperature regimes. These results demonstrate that laboratory selection or a sampling bottleneck may deplete genetic diversity at functional and possibly temperature-relevant candidate genes, but does not necessarily lead to a depletion of phenotypic variation. Furthermore, analyses described in **Chapter 3** revealed that variation in expression levels of specifically temperature-relevant candidate genes was not substantially decreased compared to the transcriptome-wide expression level variation. These observations show that the adaptive potential of *Daphnia* populations in the wild, at least regarding thermal selection, is not specifically depleted. They further support the notion that it is unlikely to observe a loss of standing genetic diversity, as observed in my study presented in **Chapter 2**, in natural *Daphnia* populations, in which effective population sizes are large and genetic diversity is conserved in the resting egg bank during temporary disturbances.

The occurrence of introgression at functional loci, as indicated in **Chapter 1**, may further contribute to the preservation of standing genetic variation in spite of increasing selective pressure. Additionally, advantageous alleles from preadapted alien genotypes may accelerate adaptation in local populations. Particularly *D. galeata* was suggested earlier to benefit from interspecific hybridization events and to consequently exhibit high plasticity and adaptive potential (Brede *et al.* 2009; Thielsch 2012), based reports of asymmetric backcrossing of interspecific hybrids into *D. galeata* (e.g., Schwenk 1993; Jankowski & Straile 2004). This perceived advantage of *D. galeata* constitutes a threat of displacement to other species if confronted with *D. galeata* immigrants and thereby a threat to biodiversity. However, several studies showed bidirectional hybridization or introgression from other species into *D. galeata*. For example, in European populations, mitochondrial introgression from *D. galeata* into *D. longispina* was observed (Brede *et al.* 2009;

Alric *et al.* 2016), as well as bidirectional hybridization between these species (although hybridization of female *D. galeata* and male *D. longispina* was more common, Griebel *et al.* 2016). In Japanese populations, nuclear introgression from *D. galeata* into *Daphnia dentifera* was reported (Ishida *et al.* 2011). Because of the fast spread of advantageous genotypes due to clonal selection in *Daphnia* during the growing season (Spitze 1991; De Meester *et al.* 2002), few hybridization events are probably sufficient to establish foreign alleles within a gene pool, provided that they offer a selective advantage. Thus, other species of the *D. longispina* complex besides *D. galeata* may also benefit from semipermeable species barriers. The assessment of the potentially asymmetric advantage of *D. galeata* would greatly benefit from information on the direction of introgression at genomic regions involved in local adaptation.

However, gene flow among populations or species may still have negative effects that possibly outweigh beneficial effects (Allendorf *et al.* 2001), since increased rates of gene flow may result in loss of biodiversity. Genetically divergent populations may become replaced by a homogenized meta-population or parental species by a hybrid swarm as observed, e.g., in several fish species (*Alosa* spec., Hasselman *et al.*, 2014; *Cobitis* spec., Kwan *et al.*, 2014; *Cyprinella* spec., Glotzbecker *et al.*, 2016) and European tree frogs (*Hyla arborea*, Dufresnes *et al.*, 2015). Based on my observations (**Chapter 1** and **3**), there is neither among *Daphnia* populations, nor among species of the *D. longispina* complex any indication for such a loss of diversity. In **Chapter 1**, I showed a clear divergence between *D. galeata* and *D. longispina* at putatively neutral genetic loci and in **Chapter 3** I demonstrated a clear clustering of *D. galeata* genotypes according to their populations based on transcriptome-wide data. This resilience against a breakdown of divergence in the *D. longispina* complex is probably explained by the buffer capacities of the resting egg banks. Since the species of this complex often do not overwinter but usually repopulate the lakes each spring from propagules (Petrusek *et al.* 2008b; Yin *et al.* 2010), resting egg banks may serve as a buffer against loss of biodiversity at two scales: Firstly, local replacement of a parental species by a homogenized hybrid-swarm is probably strongly hampered, since interspecific hybrids of *D. galeata* and *D. longispina* usually exhibit reduced fertility, i.e., produce less ephippia and have a lower hatching rate compared to the parental species (Keller *et al.* 2007). Thus, interspecific hybrids may rapidly spread within a population but usually contribute little to the resting egg bank, compared to backcrosses (which are also only produced rarely) and the parental species (Keller *et al.* 2007; Griebel *et al.* 2016); resulting in backcrossing and introgression but not in the formation of hybrid lines. Secondly, the recruitment of locally produced genotypes from the resting egg bank each spring should at least delay a replacement of genetically divergent populations by a homogenized meta-population. I therefore conclude that the buffering effect of the resting egg bank contributes to the preservation of species barriers and population divergence I observed in the *D. longispina* complex, even in disturbed environments.

Consequently, a high resilience and adaptive potential can be determined for populations of the *D. longispina* complex, despite the adverse prognosis for freshwater ectotherms confronted with advancing climate change (Gunderson & Stillman 2015). Thus far, populations of the *D. longispina* complex and specifically of *D. galeata* seem to be under no immediate danger of displacement due to increasing anthropogenic disturbance. This resilience of daphnids against increasing temperatures might furthermore contribute to the buffering capacities of freshwater ecosystems. However, since this resilience likely depends on the strength of selection pressure and the time period over which it is applied, extensive studies over time and across the species complex are needed to further confirm this conclusion. Furthermore, the direct impact of temperature is merely one threat among many in freshwater ecosystems. Species-interactions and nutrient constraints might play a more pronounced role in local adaptation than temperature, and their influence will likely increase in the future (Settele *et al.* 2014). The selection observed on *TRY5F* (**Chapter 1**) and the prominent role of proteases among gene functions associated with transcripts that contribute highly to population divergence (**Chapter 3**) might already signify responses to such alterations in species-interactions and nutrient constraints. Proteases are sensitive to cyanobacteria toxins and important in adaptation to nutrient resources (Hairston *et al.* 1999; Agrawal *et al.* 2005; Jiang *et al.* 2015), and increasing temperatures alter the composition of algal communities, favoring cyanobacteria (Paerl *et al.* 2011; Sandrini *et al.* 2016), which is possibly reflected in my results. Moreover, the situation is aggravated by threats independent from climate change, like pollution and the introduction of invasive species. Consequently, despite the positive future prospects for the species of the *D. longispina* complex regarding thermal selection, the overall outlook remains unfavorable.

Outlook

Prospects for future research

I revealed many candidate genes potentially involved in thermal adaptation in the *D. longispina* complex (**Chapter 1** and **3**) and found support for the occurrence of adaptive introgression (**Chapter 1**). These results constitute an important resource for future research on local thermal adaptation in general and in *Daphnia* populations specifically. To advance our understanding of the genetic basis of adaptation and of the role of interspecific hybridization and introgression in *Daphnia*, further analyses are necessary. Genome-wide studies through time in several admixed populations would be helpful in revealing the amount and direction of introgression as well as the affected genomic regions. Particularly the assessment of a potentially asymmetric advantage of one of the parental species of the *D. longispina* complex would greatly benefit from such information. Genome-wide data should be complemented by transcriptome-wide data, ideally across different environmental conditions. Such information would enable us to assess the interplay of sequence and expression level differences and their respective contribution to population divergence under various environmental conditions. Furthermore, such data would allow for the correction of

misassemblies and for the assessment of the role of gene duplications, alternative splicing, and non-coding regulatory elements in local adaptation. To distinguish more accurately among signals of phylogenetic, demographic and selective processes, outgroup genome, and transcriptome data would be helpful. Finally, to assess the evolutionary impact of the observed differences, it is important to include phenotypic information on fitness-related traits in analyses on the molecular basis of adaptation. Ideally, the relationship among genotypes, phenotypes, and expression profiles should be revealed by the integration of life-history experiments, and association testing, as exemplified in **Chapter 2** of this thesis. Complementarily, the comparative analysis of ecologically informative candidate genes or gene functions based on available literature (introduced in **Chapter 3**) allows for specific hypothesis testing even in absence of phenotypic information for specific species or populations. This approach depends, however, on the quality as well as on the quantity of available information on the relationship of genotypes, phenotypes, and expression profiles, and is constantly improved by additional studies on these subjects. A database on ecologically informative candidate genes across species, i.e., on genes and gene functions under selection or adaptive in specific environmental conditions, would immensely facilitate the application of the approach. It would also facilitate hypothesis testing across a broad range of organisms and environmental factors.

Closing remarks

If applied across a variety of organisms or environmental factors, the utilization of ecologically informative gene sets might reveal general patterns underlying adaptive processes and improve our understanding of evolutionary change, which is critical for informed decision-making (e.g., Whiteley *et al.* 2015). One example of this is the change of perspective in the field of conservation biology regarding interspecific hybridization: While hybrids were initially not considered worthy of protection, the beneficial effects of interspecific hybridization are more and more recognized (Allendorf *et al.* 2001; Hamilton & Miller 2016). Gene flow among populations and species potentially reduces the threat imposed by environmental disturbance, but will inevitably and permanently alter species genomes. It is now understood that the latter is not necessarily a negative effect, and that, when confronted with environmental change, selection and adaptive change are inevitable. Therefore, it is not only futile but potentially harmful for ecosystem and biodiversity protection to aim at conserving the status quo in an ever changing world.

Although immense amounts of data were gathered for a wide range of organisms, especially since the advent of high-throughput sequencing techniques, there are still more questions than answers regarding the molecular basis of adaptation. The prognosis of Holt (1990) from 27 years ago with regard to the thread imposed by climate change, "There is almost no species for which we know enough relevant ecology, physiology and genetics to predict its evolutionary response to climate change" still holds true today. Nonetheless, I am positive that methodological advances and the growing body of expression, sequence, and phenotypic data will change that in the future.

References

Abbott R, Albach D, Ansell S, *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.

Afgan E, Baker D, van den Beek M, *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* **44**, W3-W10.

Agrawal MK, Zitt A, Bagchi D, *et al.* (2005) Characterization of proteases in guts of *Daphnia magna* and their inhibition by *Microcystis aeruginosa* PCC 7806. *Environmental Toxicology* **20**, 314-322.

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

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16**, 613-622.

Alric B, Most M, Domaizon I, *et al.* (2016) Local human pressures influence gene flow in a hybridizing *Daphnia* species complex. *Journal of Evolutionary Biology* **29**, 720-735.

Altermatt F, Hottinger J, Ebert D (2007) Parasites promote host gene flow in a metapopulation. *Evolutionary Ecology* **21**, 561-575.

Andersen DH, Pertoldi C, Loeschcke V, Scali V (2006) Developmental instability, hybridization and heterozygosity in stick insects of the genus *Bacillus* (Insecta; Phasmatodea) with different modes of reproduction. *Biological Journal of the Linnean Society* **87**, 249-259.

Anderson E (1949) *Introgressive hybridization*. John Wiley & Sons, New York, USA.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217-1229.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology* **21**, 2078-2091.

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

Arnold FH, Wintrode PL, Miyazaki K, Gershenson A (2001) How enzymes adapt: lessons from directed evolution. *Trends in Biochemical Sciences* **26**, 100-106.

Arnold ML (1997) *Natural hybridization and evolution*. Oxford University Press, New York, USA.

Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution* **10**, 67-71.

Asselman J, De Coninck DIM, Glaholt S, *et al.* (2012) Identification of pathways, gene networks, and paralogous gene families in *Daphnia pulex* responding to exposure to the toxic cyanobacterium *Microcystis aeruginosa*. *Environmental Science & Technology* **46**, 8448-8457.

Asselman J, Pfrender ME, Lopez J, *et al.* (2015) Conserved transcriptional responses to cyanobacterial stressors are mediated by alternate regulation of paralogous genes in *Daphnia*. *Molecular Ecology* **24**, 1844-1855.

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

BADC (2008) Climatic Research Unit (CRU) time-series datasets of variations in climate with variations in other phenomena TS3. University of East Anglia.

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics* **4**, 99-111.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Banta AM, Wood TR, Brown LA, Ingle L (1939) *Studies on the physiology, genetics, and evolution of some Cladocera*. Carnegie Institution of Washington, Washington, D.C., USA.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* **12**, 767-780.

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11-21.

Baskett ML, Gomulkiewicz R (2011) Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology* **4**, 223-239.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Becker M, Gruenheit N, Steel M, *et al.* (2013) Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change* **3**, 1039-1043.

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

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

Berdan EL, Mazzoni CJ, Waurick I, Roehr JT, Mayer F (2015) A population genomic scan in *Chorthippus* grasshoppers unveils previously unknown phenotypic divergence. *Molecular Ecology* **24**, 3918-3930.

Bernot R, Dodds W, Quist M, Guy C (2006) Temperature and kairomone induced life history plasticity in coexisting *Daphnia*. *Aquatic Ecology* **40**, 361-372.

Berthouly-Salazar C, Thuillet A-C, Rhoné B, *et al.* (2016) Genome scan reveals selection acting on genes linked to stress response in wild pearl millet. *Molecular Ecology* **25**, 5500-5512.

Blom JF, Baumann HI, Codd GA, Juettner F (2006) Sensitivity and adaptation of aquatic organisms to oscillapeptin J and [D-Asp^3,(*E*)-Dhb^7] microcystin-RR. *Archiv Fur Hydrobiologie* **167**, 547-559.

Boersma M, De Meester L, Spaak P (1999) Environmental stress and local adaptation in *Daphnia*. *Limnology and Oceanography* **44**, 393-402.

Boersma M, Vijverberg J (1994) Seasonal variations in the condition of two *Daphnia* species and their hybrid in a eutrophic lake: evidence for food limitation. *Journal of Plankton Research* **16**, 1793-1809.

Bonhomme M, Chevalet C, Servin B, *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* **186**, 241-262.

Bradshaw AD (1965) Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics* **13**, 115-155.

Brawand D, Soumillon M, Necsulea A, *et al.* (2011) The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-348.

Brede N (2008) *The reconstruction of evolutionary patterns from* Daphnia *resting egg banks* PhD Thesis, Goethe-University.

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* **106**, 4758-4763.

Brede N, Thielsch A, Sandrock C, *et al.* (2006) Microsatellite markers for European *Daphnia*. *Molecular Ecology Notes* **6**, 536-539.

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.

Buhk C, Thielsch A (2015) Hybridisation boosts the invasion of an alien species complex: insights into future invasiveness. *Perspectives in Plant Ecology, Evolution and Systematics* **17**, 274-283.

Bürger R (2000) *The mathematical theory of selection, recombination, and mutation*. Wiley, Chichester, UK.

Caceres CE, Hairston NG (1998) Benthic-pelagic coupling in planktonic crustaceans: the role of the benthos. *Ergebnisse der Limnologie* **52**, 163-174.

Cáceres CE, Soluk DA (2002) Blowing in the wind: a field test of overland dispersal and colonization by aquatic invertebrates. *Oecologia* **131**, 402-408.

Carvalho G (1987) The clonal ecology of *Daphnia magna* (Crustacea: Cladocera): II. thermal differentiation among seasonal clones. *Journal of Animal Ecology* **56**, 469-478.

Casper P, Koschel R (1995) Description of Lake Stechlin. *Limnologica* **25**, 281-284.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* **70**, 155-174.

Charmantier A, McCleery RH, Cole LR, *et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science* **320**, 800-803.

Chaves JA, Cooper EA, Hendry AP, *et al.* (2016) Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology* **25**, 5282-5295.

Chavez-Galarza J, Henriques D, Johnston JS, *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology* **22**, 5890-5907.

Chen J, Nolte V, Schloetterer C (2015) Temperature-related reaction norms of gene expression: regulatory architecture and functional implications. *Molecular Biology and Evolution* **32**, 2393-2402.

Chopelet J, Blier PU, Dufresne F (2008) Plasticity of growth rate and metabolism in *Daphnia magna* populations from different thermal habitats. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology* **309**, 553-562.

Chunco AJ (2014) Hybridization in a warmer world. *Ecology and Evolution* **4**, 2019-2031.

Chunco AJ, Jobe T, Pfennig KS (2012) Why do species co-occur? A test of alternative hypotheses describing abiotic differences in sympatry versus allopatry using spadefoot toads. *PLoS ONE* **7**, e32748.

Colbourne JK, Pfrender ME, Gilbert D, *et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science* **331**, 555-561.

Collins M, Knutti R, Arblaster J, *et al.* (2013) Long-term climate change: projections, commitments and irreversibility. In: *Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change* eds. Stocker TF, Qin D, Plattner G-K, *et al.*), pp. 1029-1136. Cambridge University Press, Cambridge, UK.

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* **98**, 6256-6260.

Cristescu ME, Constantin A, Bock DG, Caceres CE, Crease TJ (2012) Speciation with gene flow and the genetics of habitat transitions. *Molecular Ecology* **21**, 1411-1422.

Dabney A, Storey JD (2003) Qvalue R package: Q-value estimation for false discovery rate control. Comprehensive R Archive Network.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science* **332**, 53-58.

De Coninck DIM, Asselman J, Glaholt S, *et al.* (2014) Genome-wide transcription profiles reveal genotype-dependent responses of biological pathways and gene-families in *Daphnia* exposed to single and mixed stressors. *Environmental Science & Technology* **48**, 3513-3522.

De Frenne P, Graae BJ, Rodríguez-Sánchez F, *et al.* (2013) Latitudinal gradients as natural laboratories to infer species' responses to temperature. *Journal of Ecology* **101**, 784-795.

de Jong MA, Collins S, Beldade P, Brakefield PM, Zwaan BJ (2013) Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline. *Molecular Ecology* **22**, 341-353.

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 in Ecology & Evolution* **31**, 136-146.

Decaestecker E, Gaba S, Raeymaekers JAM, *et al.* (2007) Host-parasite `Red Queen' dynamics archived in pond sediment. *Nature* **450**, 870-873.

Delmotte F, Sabater-Muñoz B, Prunier-Leterme N, *et al.* (2003) Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution* **57**, 1291-1303.

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.

Dittmar EL, Oakley CG, Conner JK, Gould BA, Schemske DW (2016) Factors influencing the effect size distribution of adaptive substitutions. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **283**, 20153065.

Dlouha S, Thielsch A, Kraus RHS, *et al.* (2010) Identifying hybridizing taxa within the *Daphnia longispina* species complex: a comparison of genetic methods and phenotypic approaches. *Hydrobiologia* **643**, 107-122.

Drummond A, Ashton B, Cheung M, Heled J, Kearse M (2011) Geneious v5.3.6. Biomatters Ltd., Auckland, New Zealand.

Du X, Li L, Zhang S, Meng F, Zhang G (2014) SNP identification by transcriptome sequencing and candidate gene-based association analysis for heat tolerance in the bay scallop *Argopecten irradians*. *PLoS ONE* **9**, e104960.

Dudgeon D, Arthington AH, Gessner MO, *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society* **81**, 163-182.

Dunn RR, Harris NC, Colwell RK, Koh LP, Sodhi NS (2009) The sixth mass coextinction: are most endangered species parasites and mutualists? *Proceedings of the Royal Society of London, Series B: Biological Sciences* **276**, 3037-3045.

Ebert D (2005) *Ecology, epidemiology, and evolution of parasitism in* Daphnia. National Center for Biotechnology Information, Bethesda (MD), USA.

Ebert D, Haag C, Kirkpatrick M, *et al.* (2002) A selective advantage to immigrant genes in a *Daphnia* metapopulation. *Science* **295**, 485-488.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

Egea R, Casillas S, Barbadilla A (2008) Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research* **36**, W157-W162.

Eierman LE, Hare MP (2016) Reef-specific patterns of gene expression plasticity in Eastern oysters (*Crassostrea virginica*). *Journal of Heredity* **107**, 90-100.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* **24**, 217-242.

Eveno E, Collada C, Guevara MA, *et al.* (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution* **25**, 417-437.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**, 610-618.

Facon B, Pointier J-P, 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.

Feldmeyer B, Mazur J, Beros S, *et al.* (2016) Gene expression patterns underlying parasite-induced alterations in host behaviour and life history. *Molecular Ecology* **25**, 648-660.

Figuerola J, Green Andy J, Michot Thomas C (2005) Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. *The American Naturalist* **165**, 274-280.

Fischer EM, Knutti R (2015) Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change* **5**, 560-564.

Fitzsimmons KE, Hambach U (2014) Loess accumulation during the last glacial maximum: evidence from Urluia, southeastern Romania. *Quaternary International* **334-335**, 74-85.

Flößner D (1972) *Krebstiere, Crustacea. Kiemen- und Blattfüsser, Branchiopoda. Fischläuse, Branchiura*. Gustav Fischer, Jena, Germany.

Flößner D, Kraus K (1986) On the taxonomy of the *Daphnia-hyalina-galeata* complex (Crustacea: Cladocera). *Hydrobiologia* **137**, 97-115.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Foran JA (1986) A comparison of the life history features of a temperate and a subtropical *Daphnia* species. *Oikos* **46**, 185-193.

Fraisse C, Belkhir K, Welch JJ, Bierne N (2016) Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology* **25**, 269-286.

Franks SJ, Kane NC, O'Hara NB, Tittes S, Rest JS (2016) Rapid genome-wide evolution in *Brassica rapa* populations following drought revealed by sequencing of ancestral and descendant gene pools. *Molecular Ecology* **25**, 3622-3631.

Franssen SU, Gu J, Winters G, *et al.* (2014) Genome-wide transcriptomic responses of the seagrasses *Zostera marina* and *Nanozostera noltii* under a simulated heatwave confirm functional types. *Marine Genomics* **15**, 65-73.

Frisch D, Morton PK, Chowdhury PR, *et al.* (2014) A millennial scale chronicle of evolutionary responses to cultural eutrophication in *Daphnia*. *Ecology Letters* **17**, 360-368.

Futuyma D (2013) *Evolution*, third edn. Sinauer Associates, Inc., Massachusets, USA.

Gademann K, Portmann C (2008) Secondary metabolites from cyanobacteria: complex structures and powerful bioactivities. *Current Organic Chemistry* **12**, 326-341.

Garroway CJ, Bowman J, Cascaden TJ, *et al.* (2010) Climate change induced hybridization in flying squirrels. *Global Change Biology* **16**, 113-121.

Gaskin JF, Wheeler GS, Purcell MF, Taylor GS (2009) Molecular evidence of hybridization in Florida's sheoak (*Casuarina* spp.) invasion. *Molecular Ecology* **18**, 3216-3226.

Gavrilets S, Gibson N (2002) Fixation probabilities in a spatially heterogeneous environment. *Population Ecology* **44**, 51-58.

Geerts AN, Vanoverbeke J, Vanschoenwinkel B, *et al.* (2015) Rapid evolution of thermal tolerance in the water flea *Daphnia*. *Nature Climate Change* **5**, 665-668.

Gibert P, De Jong G (2001) Temperature dependence of development rate and adult size in *Drosophila* species: biophysical parameters. *Journal of Evolutionary Biology* **14**, 267-276.

Gießler S, Mader E, Schwenk K (1999) Morphological evolution and genetic differentiation in *Daphnia* species complexes. *Journal of Evolutionary Biology* **12**, 710-723.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**, 242-245.

Glotzbecker GJ, Walters DM, Blum MJ (2016) Rapid movement and instability of an invasive hybrid swarm. *Evolutionary Applications* **9**, 741-755.

Gómez A, Carvalho GR (2000) Sex, parthenogenesis and genetic structure of rotifers: microsatellite analysis of contemporary and resting egg bank populations. *Molecular Ecology* **9**, 203-214.

Gómez JM, González-Megías A, Lorite J, Abdelaziz M, Perfectti F (2015) The silent extinction: climate change and the potential hybridization-mediated extinction of endemic high-mountain plants. *Biodiversity and Conservation* **24**, 1843-1857.

Gompert Z, Lucas LK, Nice CC, *et al.* (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution* **66**, 2167-2181.

Gosset CC, Bierne N (2013) Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species. *Journal of Evolutionary Biology* **26**, 14-26.

Grant BR, Grant PR (1996) High survival of Darwin's finch hybrids: effects of beak morphology and diets. *Ecology*, 500-509.

Grant PR, Grant BR (2010) Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **365**, 1065-1076.

Griebel J, Gießler S, Poxleitner M, *et al.* (2015) Extreme environments facilitate hybrid superiority ? The story of a successful *Daphnia galeata* × *longispina* hybrid clone. *PLoS ONE* **10**, e0140275.

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.

Griswold CK (2006) Gene flow's effect on the genetic architecture of a local adaptation and its consequences for QTL analyses. *Heredity* **96**, 445-453.

Gunderson AR, Stillman JH (2015) Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **282**, 20150401.

Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2005) Founder events as determinants of within-island and among-island genetic structure of *Daphnia* metapopulations. *Heredity* **96**, 150-158.

Hairston NG, Lampert W, Caceres CE, *et al.* (1999) Lake ecosystems: rapid evolution revealed by dormant eggs. *Nature* **401**, 446-446.

Hamilton JA, Lexer C, Aitken SN (2013a) Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone. *New Phytologist* **197**, 927-938.

Hamilton JA, Lexer C, Aitken SN (2013b) Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology* **22**, 827-841.

Hamilton JA, Miller JM (2016) Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* **30**, 33-41.

Hansen MM, Olivieri I, Waller DM, Nielsen EE, The GeM Working Group (2012) Monitoring adaptive genetic responses to environmental change. *Molecular Ecology* **21**, 1311-1329.

Harrison PW, Wright AE, Mank JE (2012) The evolution of gene expression and the transcriptome–phenotype relationship. *Seminars in Cell & Developmental Biology* **23**, 222-229.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity* **105**, 795-809.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology* **25**, 2454-2466.

Hasselman DJ, Argo EE, McBride MC*, et al.* (2014) Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. *Molecular Ecology* **23**, 1137-1152.

Hebert PDN (1978) The population biology of *Daphnia* (Crustacea, Daphnidae). *Biological Reviews* **53**, 387-426.

Hebert PDN, Schwartz SS, Ward RD, Finston TL (1993) Macrogeographic patterns of breeding system diversity in the *Daphnia pulex* group. I. Breeding systems of Canadian populations. *Heredity* **70**, 148-161.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.

Hedrick PW (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology* **22**, 4606-4618.

Hendry AP, Day T, Taylor EB (2001) Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution* **55**, 459-466.

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.

Henry P, Russello MA (2013) Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution* **3**, 3906-3917.

Herrmann M, Henning-Lucass N, Cordellier M, Schwenk K (in press) A genotype-phenotype association approach to reveal thermal adaptation in *Daphnia galeata*. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*.

Hoban S, Kelley JL, Lotterhos KE*, et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist* **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*, 1502-1516.

Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics* **73**, 95-108.

Hoffmann AA, Hallas RJ, Dean JA, Schiffer M (2003) Low potential for climatic stress adaptation in a rainforest *Drosophila* species. *Science* **301**, 100-102.

Hoffmann AA, Sgro CM (2011) Climate change and evolutionary adaptation. *Nature* **470**, 479-485.

Holt RD (1990) The microevolutionary consequences of climate change. *Trends in Ecology & Evolution* **5**, 311-315.

Houde M, Carter B, Douville M (2013) Sublethal effects of the flame retardant intermediate hexachlorocyclopentadiene (HCCPD) on the gene transcription and protein activity of *Daphnia magna*. *Aquatic Toxicology* **140**, 213-219.

Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153-159.

Hutchins DA, Walworth NG, Webb EA, *et al.* (2015) Irreversibly increased nitrogen fixation in *Trichodesmium* experimentally adapted to elevated carbon dioxide. *Nature Communications* **6**, 8155.

Huylmans AK, López Ezquerra A, Parsch J, Cordellier M (2016) De novo transcriptome assembly and sex-biased gene expression in the cyclical parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution* **8**, 3120-3139.

IPCC (2014) Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of Working Group II to the fifth assessment report of the Intergovernmental Panel on Climate Change eds. Field CB, Barros VR, Dokken DJ, *et al.*). Cambridge University Press, Cambridge, UK and New York, USA.

Ishida S, Takahashi A, Matsushima N, *et al.* (2011) The long-term consequences of hybridization between the two *Daphnia* species, *D. galeata* and *D. dentifera*, in mature habitats. *BMC Evolutionary Biology* **11**, 209.

Ishida S, Taylor DJ (2007a) Mature habitats associated with genetic divergence despite strong dispersal ability in an arthropod. *BMC Evolutionary Biology* **7**, 52.

Ishida S, Taylor DJ (2007b) Quaternary diversification in a sexual Holarctic zooplankter, *Daphnia galeata*. *Molecular Ecology* **16**, 569-582.

James AC, Azevedo RBR, Partridge L (1997) Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. *Genetics* **146**, 881-890.

Jankowski T, Straile D (2003) A comparison of egg-bank and long-term plankton dynamics of two *Daphnia* species, *D. hyalina* and *D. galeata*: potentials and limits of reconstruction. *Limnology and Oceanography* **48**, 1948-1955.

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

Jansen M, Geerts AN, Rago A, *et al.* (2017) Thermal tolerance in the keystone species *Daphnia magna* –a candidate gene and an outlier analysis approach. *Molecular Ecology* **26**, 2291-2305.

Jensen JD, Foll M, Bernatchez L (2016) The past, present and future of genomic scans for selection. *Molecular Ecology* **25**, 1-4.

Jeyasingh PD, Ragavendran A, Paland S, *et al.* (2011) How do consumers deal with stoichiometric constraints? Lessons from functional genomics using *Daphnia pulex*. *Molecular Ecology* **20**, 2341-2352.

Ji X, Qiu Q-B, Diong C-H (2002) Influence of incubation temperature on hatching success, energy expenditure for embryonic development, and size and morphology of hatchlings in the oriental garden lizard, *Calotes versicolor* (Agamidae). *Journal of Experimental Zoology* **292**, 649-659.

Jiang X, Liang H, Chen Y, Xu X, Huang D (2015) Microgeographic adaptation to toxic cyanobacteria in two aquatic grazers. *Limnology and Oceanography* **60**, 947-956.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 1-15.

Joost S, Vuilleumier S, Jensen JD, *et al.* (2013) Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Molecular Ecology* **22**, 3659-3665.

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.

Kanshin E, Kubiniok P, Thattikota Y, D'Amours D, Thibault P (2015) Phosphoproteome dynamics of *Saccharomyces cerevisiae* under heat shock and cold stress. *Molecular Systems Biology* **11**, 813.

Karl I, Schmitt T, Fischer K (2008) Phosphoglucose isomerase genotype affects life history traits and cold stress resistance in a copper butterfly. *Functional Ecology* **22**, 887-894.

Kearsey MJ (1998) The principles of QTL analysis (a minimal mathematics approach). *Journal of Experimental Botany* **49**, 1619-1623.

Keller B, Spaak P (2004) Nonrandom sexual reproduction and diapausing egg production in a *Daphnia* hybrid species complex. *Limnology and Oceanography* **49**, 1393-1400.

Keller B, Wolinska J, Manca M, Spaak P (2008) Spatial, environmental and anthropogenic effects on the taxon composition of hybridizing *Daphnia*. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **363**, 2943-2952.

Keller B, Wolinska J, Tellenbach C, Spaak P (2007) Reproductive isolation keeps hybridizing *Daphnia* species distinct. *Limnology and Oceanography* **52**, 984-991.

Kelly MW, Sanford E, Grosberg RK (2012a) Limited potential for adaptation to climate change in a broadly distributed marine crustacean. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **279**, 349-356.

Kelly SA, Panhuis TM, Stoehr AM (2012b) Phenotypic plasticity: molecular mechanisms and adaptive significance. *Compr Physiol* **2**, 1417-1439.

Kerfoot CW, Ma X, Lorence CS, Weider LJ (2004) Toward resurrection ecology: *Daphnia mendotae* and *D. retrocurva* in the coastal region of Lake Superior, among the first successful outside invaders? *Journal of Great Lakes Research* **30 (Supplement 1)**, 285-299.

Kerfoot WC (1974) Net accumulation rates and the history of cladoceran communities. *Ecology* **55**, 51-61.

Kerfoot WC, Robbins JA, Weider LJ (1999) A new approach to historical reconstruction: combining descriptive and experimental paleolimnology. *Limnology and Oceanography* **44**, 1232-1247.

Khan AH, Lin A, Smith DJ (2012) Discovery and characterization of human exonic transcriptional regulatory elements. *PLoS ONE* **7**, e46098.

Klüttgen B, Dülmer U, Engels M, Ratte HT (1994) ADaM, an artificial freshwater for the culture of zooplankton. *Water Research* **28**, 743-746.

Kolbe JJ, Glor RE, Rodriguez Schettino L*, et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature* **431**, 177-181.

Kopp M, Hermisson J (2007) Adaptation of a quantitative trait to a moving optimum. *Genetics* **176**, 715-719.

Kopp M, Hermisson J (2009) The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics* **183**, 1453-1476.

Korkmaz S, Goksuluk D, Zararsiz G (2014) MVN: An R package for assessing multivariate normality. *R Journal* **6**, 151-162.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**, 29-29.

Kristensen TN, Overgaard J, Lassen J, Hoffmann AA, Sgro C (2015) Low evolutionary potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures. *Evolution* **69**, 803-814.

Kubecka J, Bohm M (1991) The fish fauna of the Jordan reservoir, one of the oldest man-made lakes in central Europe. *Journal of Fish Biology* **38**, 935-950.

Kwan YS, Ko MH, Won YJ (2014) Genomic replacement of native *Cobitis lutheri* with introduced *C. tetralineata* through a hybrid swarm following the artificial connection of river systems. *Ecology and Evolution* **4**, 1451-1465.

Lampert W (2011) Daphnia*: development of a model organism in ecology and evolution*. International Ecology Institute, Oldendorf/Luhe, Germany.

Lampert W, Trubetskova I (1996) Juvenile growth rate as a measure of fitness in *Daphnia*. *Functional Ecology* **10**, 631-635.

Lande R (1988) Genetics and demography in biological conservation. *Science* **241**, 1455-1460.

Landis G, Shen J, Tower J (2012) Gene expression changes in response to aging compared to heat stress, oxidative stress and ionizing radiation in *Drosophila melanogaster*. *Aging-Us* **4**, 768-789.

Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013) Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution* **67**, 3653-3661.

Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of species boundaries. *Molecular Ecology* **23**, 1668-1678.

Latta LC, Weider LJ, Colbourne JK, Pfrender ME (2012) The evolution of salinity tolerance in *Daphnia*: a functional genomics approach. *Ecology Letters* **15**, 794-802.

Laurent S, Pfeifer SP, Settles ML, *et al.* (2016) The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. *Molecular Ecology* **25**, 306-323.

Le Gac M, Metegnier G, Chomerat N, *et al.* (2016) Evolutionary processes and cellular functions underlying divergence in *Alexandrium minutum*. *Molecular Ecology* **25**, 5129-5143.

Lennon JT, Smith VH, Williams K (2001) Influence of temperature on exotic *Daphnia lumholtzi* and implications for invasion success. *Journal of Plankton Research* **23**, 425-433.

Levine MT, Eckert ML, Begun DJ (2011) Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from Eastern Australia. *Molecular Biology and Evolution* **28**, 249-256.

Lewontin RC, Birch LC (1966) Hybridization as a source of variation for adaptation to new environments. *Evolution* **20**, 315-336.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-2189.

Li YF, Costello JC, Holloway AK, Hahn MW, Rausher M (2008) "Reverse ecology" and the power of population genomics. *Evolution* **62**, 2984-2994.

Li Z, Hu G, Liu X, *et al.* (2016) Transcriptome sequencing identified genes and gene ontologies associated with early freezing tolerance in maize. *Frontiers in Plant Science* **7**, 1477.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Liu H, Lamm MS, Rutherford K, *et al.* (2015) Large-scale transcriptome sequencing reveals novel expression patterns for key sex-related genes in a sex-changing fish. *Biology of Sex Differences* **6**, 26.

Livingstone DM, Adrianb R (2009) Modeling the duration of intermittent ice cover on a lake for climate-change studies. *Limnology and Oceanography* **54**, 1709-1722.

Livingstone DM, Lotter AF (1998) The relationship between air and water temperatures in lakes of the Swiss Plateau: a case study with palaeolimnological implications. *Journal of Paleolimnology* **19**, 181-198.

Lloyd MM, Makukhov AD, Pespeni MH (2016) Loss of genetic diversity as a consequence of selection in response to high pCO_2. *Evolutionary Applications* **9**, 1124-1132.

Lohr JN, Haag CR (2015) Genetic load, inbreeding depression, and hybrid vigor covary with population size: An empirical evaluation of theoretical predictions. *Evolution* **69**, 3109-3122.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23**, 2178-2192.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Lucek K, Lemoine M, Seehausen O (2014) Contemporary ecotypic divergence during a recent range expansion was facilitated by adaptive introgression. *Journal of Evolutionary Biology* **27**, 2233-2248.

Lucek KAY, Roy D, Bezault E, Sivasundar A, Seehausen OLE (2010) Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. *Molecular Ecology* **19**, 3995-4011.

Luttikhuizen PC, Drent J, Peijnenburg KT, van der Veer HW, Johannesson K (2012) Genetic architecture in a marine hybrid zone: comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*. *Molecular Ecology* **21**, 3048-3061.

Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, 303-339.

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237.

Manel S, Joost S, Epperson BK, *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760-3772.

Manríquez-Morán NL, Cruz FRM-dl, Murphy RW (2014) Genetic variation and origin of parthenogenesis in the *Aspidoscelis cozumela* complex: evidence from mitochondrial genes. *Zoological Science* **31**, 14-19.

Markova S, Dufresne F, Manca M, Kotlik P (2013) Mitochondrial capture misleads about ecological speciation in the *Daphnia pulex* complex. *PLoS ONE* **8**, e69497.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology* **24**, 3009-3025.

Matuszewski S, Hermisson J, Kopp M (2014) Fisher's geometric model with a moving optimum. *Evolution* **68**, 2571-2588.

Mayr E (1942) *Systematics and the origin of species*. Columbia University Press, New York, USA.

Mayr E (1963) *Animal species and evolution*. Harvard University Press, Cambridge, UK.

Mazdiyasni O, AghaKouchak A (2015) Substantial increase in concurrent droughts and heatwaves in the United States. *Proceedings of the National Academy of Sciences* **112**, 11484-11489.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Mergeay J, Verschuren D, De Meester L (2005) *Daphnia* species diversity in Kenya, and a key to the identification of their ephippia. *Hydrobiologia* **542**, 261-274.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications* **7**, 1-14.

Michels H, Amsinck SL, Jeppesen E, De Meester L (2007) Interclonal variation in diel horizontal migration behaviour of the water flea *Daphnia magna*—searching for a signature of adaptive evolution. *Hydrobiologia* **594**, 117-129.

Millennium Ecosystem Assessment (2005) Ecosystem services and human well-being: Wetlands & Water: Synthesis. A Report of the Millennium Ecosystem Assessment. World Resources Institute, Washington, D.C., USA.

Mitchell SE, Lampert W (2000) Temperature adaptation in a geographically widespread zooplankter, *Daphnia magna*. *Journal of Evolutionary Biology* **13**, 371-382.

Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**, 693-712.

Montero-Pau J, Gomez A, Munoz J (2008) Application of an inexpensive and high-throughput genomic DNA extraction method for the molecular ecology of zooplanktonic diapausing eggs. *Limnology and Oceanography: Methods* **6**, 218-222.

Nolte C, Staiger D (2015) RNA around the clock - regulation at the RNA level in biological timing. *Frontiers in Plant Science* **6**, 311.

Norris LC, Main BJ, Lee Y, *et al.* (2015) Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences* **112**, 815-820.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

Orcutt JD, Porter KG (1983) Diel vertical migration by zooplankton: constant and fluctuating temperature effects on life history parameters of *Daphnia*. *Limnology and Oceanography* **28**, 720-730.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**, 119-127.

Orsini L, Schwenk K, De Meester L, *et al.* (2013) The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology & Evolution* **28**, 274-282.

Orsini L, Spanier KI, De Meester L (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: validation in space, time and experimental evolution. *Molecular Ecology* **21**, 2160-2175.

Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends in Genetics* **29**, 66-73.

Padfield D, Yvon-Durocher G, Buckling A, Jennings S, Yvon-Durocher G (2016) Rapid evolution of metabolic traits explains thermal adaptation in phytoplankton. *Ecology Letters* **19**, 133-142.

Paerl HW, Hall NS, Calandrino ES (2011) Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment* **409**, 1739-1745.

Paerl HW, Huisman J (2008) Blooms like it hot. *Science* **320**, 57-58.

Pai AA, Pritchard JK, Gilad Y (2015) The genetic and mechanistic basis for variation in gene regulation. *PLoS Genetics* **11**, e1004857.

Palsson S (2000) Microsatellite variation in *Daphnia pulex* from both sides of the Baltic Sea. *Molecular Ecology* **9**, 1075-1088.

Parchman TL, Gompert Z, Braun MJ, *et al.* (2013) The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular Ecology* **22**, 3304-3317.

Pardo-Diaz C, Salazar C, Baxter SW, *et al.* (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics* **8**, e1002752.

Pardo Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution* **6**, 445-464.

Pariset L, Joost S, Marsan PA, Valentini A, Econogene C (2009) Landscape genomics and biased F_{ST} approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics* **10**, 7.

Parris MJ (2001) High larval performance of leopard frog hybrids: effects of environment-dependent selection. *Ecology* **82**, 3001-3009.

Paul RJ, Lamkemeyer T, Maurer J, *et al.* (2004a) Thermal acclimation in the microcrustacean *Daphnia*: a survey of behavioural, physiological and biochemical mechanisms. *Journal of Thermal Biology* **29**, 655-662.

Paul RJ, Zeis B, Lamkemeyer T, Seidl M, Pirow R (2004b) Control of oxygen transport in the microcrustacean *Daphnia*: regulation of haemoglobin expression as central mechanism of adaptation to different oxygen and temperature conditions. *Acta Physiologica Scandinavica* **182**, 259-275.

Pavey SA, Collin H, Nosil P, Rogers SM (2010) The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences* **1206**, 110-129.

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Peñuelas J, Sardans J, Estiarte M, *et al.* (2013) Evidence of current impact of climate change on life: a walk from genes to the biosphere. *Global Change Biology* **19**, 2303-2338.

Pereira RJ, Barreto FS, Pierce NT, Carneiro M, Burton RS (2016) Transcriptome-wide patterns of divergence during allopatric evolution. *Molecular Ecology* **25**, 1478-1493.

Petrusek A, Hobæk A, Nilssen JP, *et al.* (2008a) A taxonomic reappraisal of the European *Daphnia longispina* complex (Crustacea, Cladocera, Anomopoda). *Zoologica Scripta* **37**, 507-519.

Petrusek A, Seda J, Machá ek J, Ruthová Š, Šmilauer P (2008b) *Daphnia* hybridization along ecological gradients in pelagic environments: the potential for the presence of hybrid zones in plankton. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **363**, 2931-2941.

Petrusek A, Thielsch A, Schwenk K (2012) Mitochondrial sequence variation suggests extensive cryptic diversity within the Western Palaearctic *Daphnia longispina* complex. *Limnology and Oceanography* **57**, 1838-1845.

Pörtner HO (2001) Climate change and temperature-dependent biogeography: oxygen limitation of thermal tolerance in animals. *Naturwissenschaften* **88**, 137-146.

Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.

R Core Team (2014) R: a Language and environment for statistical computing. R Foundation for Statistical Computing.

Raeymaekers JAM, Van Houdt JKJ, Larmuseau MHD, Geldof S, Volckaert FAM (2007) Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Molecular Ecology* **16**, 891-905.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**, 4348-4370.

Rellstab C, Keller B, Girardclos S, Anselmetti FS, Spaak P (2011) Anthropogenic eutrophication shapes the past and present taxonomic composition of hybridizing *Daphnia* in unproductive lakes. *Limnology and Oceanography* **56**, 292-302.

Renzi F, Caffarelli E, Laneve P, *et al.* (2006) The structure of the endoribonuclease XendoU: from small nucleolar RNA processing to severe acute respiratory syndrome coronavirus replication. *Proceedings of the National Academy of Sciences* **103**, 12365-12370.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* **27**, 83-109.

Rice P, Longden I, Bleasby A (2000) EMBOSS: the european molecular biology open software suite. *Trends in Genetics* **16**, 276-277.

Richards CL, Hanzawa Y, Katari MS, *et al.* (2009) Perspectives on ecological and evolutionary systems biology. *Annual Plant Reviews* **35**, 331-349.

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity* **83**, 363-372.

Rieseberg LH, Carney SE (1998) Plant hybridization. *New Phytologist* **140**, 599-624.

Rieseberg LH, Raymond O, Rosenthal DM, *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211-1216.

Rinke K, Petzoldt T (2003) Modelling the effects of temperature and food on individual growth and reproduction of *Daphnia* and their consequences on the population level. *Limnologica* **33**, 293-304.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1-17.

Roffler GH, Amish SJ, Smith S, *et al.* (2016) SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources* **16**, 1147-1164.

Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* **13**, 505-516.

Roulin AC, Routtu J, Hall MD, *et al.* (2013) Local adaptation of sex induction in a facultative sexual crustacean: insights from QTL mapping and natural populations of *Daphnia magna*. *Molecular Ecology* **22**, 3567-3579.

Salamin N, Wüest RO, Lavergne S, Thuiller W, Pearman PB (2010) Assessing rapid evolution in a changing environment. *Trends in Ecology & Evolution* **25**, 692-698.

Sandrini G, Ji X, Verspagen JMH, *et al.* (2016) Rapid adaptation of harmful cyanobacteria to rising CO_2. *Proceedings of the National Academy of Sciences* **113**, 9315-9320.

Sarnelle O, Wilson AE (2005) Local adaptation of *Daphnia pulicaria* to toxic cyanobacteria. *Limnology and Oceanography* **50**, 1565-1570.

Schneider SH, Semenov S, Patwardhan A, *et al.* (2007) Assessing key vulnerabilities and the risk from climate change. In: *Climate change 2007: impacts, adaptation, and vulnerability. Contribution of Working Group II to the fouth assessment report of the Intergovernmental Panel on Climate Change* eds. Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden PJ, Hanson, C.E.), pp. 779-810. Intergovernmental Panel on Climate Change, Cambridge, UK.

Schoville S, Barreto F, Moy G, Wolff A, Burton R (2012) Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*. *BMC Evolutionary Biology* **12**, 170.

Schumpert C, Handy I, Dudycha JL, Patel RC (2014) Relationship between heat shock protein 70 expression and life span in *Daphnia*. *Mechanisms of Ageing and Development* **139**, 1-10.

Schwarzenberger A, Kuster CJ, Elert E (2012) Molecular mechanisms of tolerance to cyanobacterial protease inhibitors revealed by clonal differences in *Daphnia magna*. *Molecular Ecology* **21**, 4898-4911.

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. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **267**, 1833-1842.

Schwenk K, Spaak P (1995) Evolutionary and ecological consequences of interspecific hybridization in cladocerans. *Experientia* **51**, 465-481.

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 °C) mainly affects protein metabolism. *BMC Physiology* **9**, 8.

Sedghifar A, Saelao P, Begun DJ (2016) Genomic patterns of geographic differentiation in *Drosophila simulans*. *Genetics* **202**, 1229-1240.

Sedlazeck FJ, Rescheneder P, von Haeseler A (2013) NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790-2791.

Seebacher F, Brand MD, Else PL, *et al.* (2010) Plasticity of oxidative metabolism in variable climates: molecular mechanisms. *Physiological and Biochemical Zoology* **83**, 721-732.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**, 198-207.

Seehausen OLE, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology* **17**, 30-44.

Seidendorf B, Boersma M, Schwenk K (2007) Evolutionary stoichiometry: the role of food quality for clonal differentiation and hybrid maintenance in a *Daphnia species* complex. *Limnology and Oceanography* **52**, 385-394.

Seidl MD, Pirow R, Paul RJ (2005) Acclimation of the microcrustacean *Daphnia magna* to warm temperatures is dependent on haemoglobin expression. *Journal of Thermal Biology* **30**, 532-544.

Settele J, Scholes R, Betts RA, *et al.* (2014) Terrestrial and inland water systems. In: *Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of Working Group II to the fifth assessment report of the Intergovernmental Panel on Climate Change* eds. Field CB, Barros VR, Dokken DJ, *et al.*), pp. 271-359. Cambridge University Press, Cambridge, UK and New York, USA.

Shapiro MD, Marks ME, Peichel CL, *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717-723.

Shikano T, Ramadevi J, Merilä J (2010) Identification of local- and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution* **27**, 2775-2789.

Shimada Y, Shikano T, Merila J (2011) A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution* **28**, 181-193.

Shulman MJ, Ogden JC, Ebersole JP, *et al.* (1983) Priority effects in the recruitment of juvenile coral reef fishes. *Ecology* **64**, 1508-1513.

Simoniello P, Esposito MG, Trinchella F, Motta CM, Scudiero R (2016) Alterations in brain morphology and HSP70 expression in lizard embryos exposed to thermal stress. *Comptes Rendus Biologies* **339**, 380-390.

Smith NG, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022-1024.

Song Y, Endepols S, Klemann N, *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Current Biology* **21**, 1296-1301.

Sørensen JG, Nielsen MM, Loeschcke V (2007) Gene expression profile analysis of *Drosophila melanogaster* selected for resistance to environmental stressors. *Journal of Evolutionary Biology* **20**, 1624-1636.

Spaak P, Fox J, Hairston NG (2012) Modes and mechanisms of a *Daphnia* invasion. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **279**, 2936-2944.

Spitze K (1991) *Chaoborus* predation and life-history evolution in *Daphnia pulex*: temporal pattern of population diversity, fitness, and mean life history. *Evolution* **45**, 82-92.

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics* **135**, 367-374.

Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* **8**, e65245.

Sterner RW, Elser JJ, Hessen DO (1992) Stoichiometric relationships among producers, consumers and nutrient cycling in pelagic ecosystems. *Biogeochemistry* **17**, 49-67.

Stoks R, Geerts AN, De Meester L (2014) Evolutionary and plastic responses of freshwater invertebrates to climate change: realized patterns and future potential. *Evolutionary Applications* **7**, 42-55.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Stuglik MT, Babik W (2016) Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecology and Evolution* **6**, 4513-4525.

Svendsen JI, Alexanderson H, Astakhov VI, *et al.* (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews* **23**, 1229-1271.

Szulkin M, Gagnaire PA, Bierne N, Charmantier A (2016) Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Molecular Ecology* **25**, 542-558.

Taylor DJ, Hebert PD (1993) Habitat-dependent hybrid parentage and differential introgression between neighboringly sympatric *Daphnia* species. *Proceedings of the National Academy of Sciences* **90**, 7079-7083.

Taylor DJ, Hebert PD, 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.

Taylor DJ, Sprenger HL, Ishida S (2005) Geographic and phylogenetic evidence for dispersed nuclear introgression in a daphniid with sexual propagules. *Molecular Ecology* **14**, 525-537.

Taylor SA, Curry RL, White TA, Ferretti V, Lovette I (2014) Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution* **68**, 3066-3081.

Teets NM, Denlinger DL (2016) Quantitative phosphoproteomics reveals signaling mechanisms associated with rapid cold hardening in a chill-tolerant fly. *Journal of Proteome Research* **15**, 2855-2862.

Teplitsky C, Mills JA, Alho JS, Yarrall JW, Merilä J (2008) Bergmann's rule and climate change revisited: disentangling environmental and genetic responses in a wild bird population. *Proceedings of the National Academy of Sciences* **105**, 13492-13496.

The UniProt Consortium (2014) UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204-D212.

Thielsch A (2012) *Population genetic structure in European* Hyalodaphnia *species: monopolization versus gene flow* PhD Thesis, Universität Koblenz-Landau.

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, Voelker E, Kraus RHS, Schwenk K (2012) Discrimination of hybrid classes using cross-species amplification of microsatellite loci: methodological challenges and solutions in *Daphnia*. *Molecular Ecology Resources* **12**, 697-705.

Todd EV, Black MA, Gemmell NJ (2016) The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology* **25**, 1224-1241.

Toews DPL, Campagna L, Taylor SA, *et al.* (2016) Genomic approaches to understanding population divergence and speciation in birds. *The Auk* **133**, 13-30.

Turgeon J, Tayeh A, Facon B, *et al.* (2011) Experimental evidence for the phenotypic impact of admixture between wild and biocontrol Asian ladybird (*Harmonia axyridis*) involved in the European invasion. *Journal of Evolutionary Biology* **24**, 1044-1052.

Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* **179**, 455-473.

Vadadi-Fülöp C, Sipkay C, Mészáros G, Hufnagel L (2012) Climate change and freshwater zooplankton: what does it boil down to? *Aquatic Ecology* **46**, 501-519.

Van Doorslaer W, Stoks R, Duvivier C, Bednarska A, De Meester L (2009a) Population dynamics determine genetic adaptation to temperature in *Daphnia*. *Evolution* **63**, 1867-1878.

Van Doorslaer W, Stoks R, Swillen I, *et al.* (2010) Experimental thermal microevolution in community-embedded *Daphnia* populations. *Climate Research* **43**, 81-89.

Van Doorslaer W, Vanoverbeke J, Duvivier C, *et al.* (2009b) Local adaptation to higher temperatures reduces immigration success of genotypes from a warmer region in the water flea *Daphnia*. *Global Change Biology* **15**, 3046-3055.

Van Gremberghe I, Vanormelingen P, Van der Gucht K, *et al.* (2009) Priority effects in experimental populations of the cyanobacterium *Microcystis*. *Environmental Microbiology* **11**, 2564-2573.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

Vandekerkhove J, Declerck S, Brendonck L, *et al.* (2005) Hatching of cladoceran resting eggs: temperature and photoperiod. *Freshwater Biology* **50**, 96-104.

Vanoverbeke J, De Meester L (1997) 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.

Velculescu VE, Zhang L, Zhou W, *et al.* (1997) Characterization of the yeast transcriptome. *Cell* **88**, 243-251.

Vera M, Diez-del-Molino D, Garcia-Marin JL (2016) Genomic survey provides insights into the evolutionary changes that occurred during European expansion of the invasive mosquitofish (*Gambusia holbrooki*). *Molecular Ecology* **25**, 1089-1105.

Vermeulen CJ, Sorensen P, Gagalova KK, Loeschcke V (2014) Flies who cannot take the heat: genome-wide gene expression analysis of temperature-sensitive lethality in an inbred line of *Drosophila melanogaster*. *Journal of Evolutionary Biology* **27**, 2152-2162.

Vermeulen CJ, Sørensen P, Kirilova Gagalova K, Loeschcke V (2013) Transcriptomic analysis of inbreeding depression in cold-sensitive *Drosophila melanogaster* shows upregulation of the immune response. *Journal of Evolutionary Biology* **26**, 1890-1902.

Vernot B, Akey JM (2014) Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017-1021.

von Elert E, Agrawal MK, Gebauer C, *et al.* (2004) Protease activity in gut of *Daphnia magna*: evidence for trypsin and chymotrypsin enzymes. *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* **137**, 287-296.

von Elert E, Zitt A, Schwarzenberger A (2012) Inducible tolerance to dietary protease inhibitors in *Daphnia magna*. *The Journal of Experimental Biology* **215**, 2051-2059.

von Heckel K, Stephan W, Hutter S (2016) Canalization of gene expression is a major signature of regulatory cold adaptation in temperate *Drosophila melanogaster*. *BMC Genomics* **17**.

Vuilleumier S, Goudet J, Perrin N (2010) Evolution in heterogeneous populations: from migration models to fixation probabilities. *Theoretical Population Biology* **78**, 250-258.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.

Weider LJ, Wolf HG (1991) Life-history variation in a hybrid species complex of *Daphnia*. *Oecologia* **87**, 506-513.

Weyhenmeyer GA, Livingstone DM, Meili M, *et al.* (2011) Large geographical differences in the sensitivity of ice-covered lakes and rivers in the Northern Hemisphere to temperature changes. *Global Change Biology* **17**, 268-275.

Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences* **103**, 5425-5430.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. *Trends in Ecology & Evolution* **30**, 42-49.

Whitman D, Agrawal A (2009) What is phenotypic plasticity and why is it important? In: *Phenotypic Plasticity of Insects - Mechanisms and Consequences* (eds. Whitman D, Ananthakrishnan T), pp. 1–63. CRC Press, Boca Raton, USA.

Whitney Kenneth D, Randell Rebecca A, Rieseberg Loren H (2006) Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* **167**, 794-807.

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics* **37**, 433-458.

Williams P, Dick K, Yampolsky L (2012) Heat tolerance, temperature acclimation, acute oxidative damage and canalization of haemoglobin expression in *Daphnia*. *Evolutionary Ecology* **26**, 591-609.

Winder M, Schindler DE (2004) Climatic effects on the phenology of lake processes. *Global Change Biology* **10**, 1844-1856.

Wojtal-Frankiewicz A (2012) The effects of global warming on *Daphnia* spp. population dynamics: a review. *Aquatic Ecology* **46**, 37-53.

Wood JLA, Yates MC, Fraser DJ (2016) Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications* **9**, 640-657.

Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics* **8**, 206-216.

Wray GA, Hahn MW, Abouheif E, *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20**, 1377-1419.

Wright S (1942) Statistical genetics and evolution. *Bulletin of the American Mathematical Society* **48**, 223-246.

Xu Q, Xing SL, Zhu CY, *et al.* (2015) Population transcriptomics reveals a potentially positive role of expression diversity in adaptation. *Journal of Integrative Plant Biology* **57**, 284-299.

Yampolsky LY, Schaer TMM, Ebert D (2014a) Adaptive phenotypic plasticity and local adaptation for temperature tolerance in freshwater zooplankton. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **281**, 20132744.

Yampolsky LY, Zeng EL, Lopez J, *et al.* (2014b) Functional genomics of acclimation and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics* **15**, 859.

Yang J, Li WR, Lv FH, *et al.* (2016a) Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular Biology and Evolution* **33**, 2576-2592.

Yang WZ, Qi Y, Fu JZ (2016b) Genetic signals of high-altitude adaptation in amphibians: a comparative transcriptome analysis. *BMC Genetics* **17**.

Ye Z, Xu S, Spitze K, *et al.* (2017) A new reference genome assembly for the microcrustacean *Daphnia pulex*. *G3: Genes | Genomes | Genetics*, g3-116.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897-1911.

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**, 80.

Yin MB, Wolinska J, Gieler 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.

Yurista PM (1999) Temperature-dependent energy budget of an Arctic cladoceran, *Daphnia middendorffiana*. *Freshwater Biology* **42**, 21-34.

Zeis B, Becker D, Gerke P, Koch M, Paul RJ (2013) Hypoxia-inducible haemoglobins of *Daphnia pulex* and their role in the response to acute and chronic temperature increase. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1834**, 1704-1710.

Zhao L, Wit J, Svetec N, Begun DJ (2015) Parallel gene expression differences between low and high latitude populations of *Drosophila melanogaster* and *D. simulans*. *PLoS Genetics* **11**, e1005184.

Zhou S, Campbell TG, Stone EA, Mackay TF, Anholt RR (2012) Phenotypic plasticity of the *Drosophila* transcriptome. *PLoS Genetics* **8**, e1002593.

Zielinski P, Dudek K, Stuglik MT, Liana M, Babik W (2014) Single nucleotide polymorphisms reveal genetic structuring of the carpathian newt and provide evidence of interspecific gene flow in the nuclear genome. *PLoS ONE* **9**, e97431.

Author contributions

The work presented in **Chapter 1** was submitted for publication in BMC Evolutionary Biology as *Natural selection in* Daphnia *populations across time, space and species boundaries* by Maike Herrmann (author of this declaration), Mathilde Cordellier (secondary thesis supervisor), Pia Kreuzer, and Klaus Schwenk (primary thesis supervisor). This study was conceived by Klaus Schwenk and me. I developed the candidate gene markers, conducted the candidate gene sequencing and data analyses and wrote the manuscript.

The work presented in **Chapter 2** is accepted for publication in the Journal of Experimental Zoology Part A: Ecological Genetics and Physiology as *A genotype-phenotype association approach to reveal thermal adaptation in* Daphnia galeata by Maike Herrmann (author of this declaration), Nicole Henning-Lucass, Mathilde Cordellier (secondary thesis supervisor) and Klaus Schwenk (primary thesis supervisor). This study was conceived by Klaus Schwenk and me. Nicole Henning-Lucass and I wrote the manuscript. I carried out or supervised genotyping of *Daphnia* clones and conducted analyses of amplicon-sequence data and association analyses.

The work presented in **Chapter 3** was submitted for publication in Molecular Ecology as *Population transcriptomics in* Daphnia*: the role of thermal selection* by Maike Herrmann (author of this declaration), Suda Parimala Ravindran, Klaus Schwenk (primary thesis supervisor) and Mathilde Cordellier (secondary thesis supervisor). This study was conceived by Mathilde Cordellier and me. I conducted the statistical data analyses and wrote the manuscript.

Curriculum vitae

Personal data

Contact	Maike Herrmann
	[redacted]
Nationality	German
Phone	[redacted]
e-mail	[redacted]

Employment

June 2015 - December 2016	Research assistant and PhD fellowship in the group of Prof. Klaus Schwenk at the Universität Koblenz-Landau
February 2012 - January 2015	PhD fellowship at the Goethe University in Frankfurt am Main and the Biodiversity and Climate Research Centre (Bik-F) in Frankfurt am Main
August 2011 - October 2011	Graduate research assistant in the group of Prof. Klaus Schwenk at the Goethe University in Frankfurt am Main
February 2011 - June 2011	Graduate research assistant in the group of Prof. Klaus Schwenk at Goethe University in Frankfurt am Main
October 2008 - May 2010	Student research assistant in the group of Prof. Klaus Schwenk at Goethe University in Frankfurt am Main

Education

since January 2015	PhD student at the Universität Koblenz-Landau the Biodiversity and the Climate Research Centre (Bik-F) in Frankfurt am Main
2012 - 2015	PhD student at the Goethe University in Frankfurt am Main and the Climate Research Centre (Bik-F) in Frankfurt am Main (Senckenberg stipend granted by the Senckenberg Gesellschaft für Naturforschung)
2005 - 2010	Diploma in Biological Sciences, Goethe University, Frankfurt am Main
	Diploma thesis: „Genetische Variabilität von natürlichen *Daphnia*-Populationen in Raum und Zeit" ("Genetic variability of natural *Daphnia* populations in space and time")

September 2005 - November 2010	Studies of Biological Sciences, Goethe University, Frankfurt am Main Main subjects: Ecology and Evolution of Animals, Ecology and Evolution of Plants, and Genetics

Scientific contributions

Publications (peer-reviewed)

***M. Herrmann**, N. Henning-Lucass, M. Cordellier, K. Schwenk* (in press): A genotype-phenotype association approach to reveal thermal adaptation in *Daphnia galeata*. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology. DOI: 10.1002/jez.2070

Talks

"Ecological Genomics in *Daphnia*", The 20th European Meeting of PhD Students in Evolutionary Biology (EMPSEB) 2014, La-Rôche-sur-Ardenne, Belgium

"Adaptive introgression in natural *Daphnia* populations", The 47th Population Genetics Group Meeting 2014, Bath, UK

"Adaptive introgression in natural *Daphnia* populations", The Daphnia Genomics Consortium Meeting 2014, Birmingham, UK

"Genetic variability of natural *Daphnia* populations in space and time", The 18th European Meeting of PhD Students in Evolutionary Biology (EMPSEB) 2012, Virrat, Finland

Poster

"Adaptive introgression in natural *Daphnia* populations" ***M. Herrmann**, P. Kreuzer, M. Cordellier, K. Schwenk,* 10th Symposium on Cladocera 2014, Lednice, Czech Republic

"Genetic basis of local and temporal adaptation in natural *Daphnia* populations" ***M. Herrmann**, P. Kreuzer, M. Cordellier, K. Schwenk*, XIV Congress of the European Society for Evolutionary Biology (ESEB) 2013, Lisbon, Portugal

"Genetic basis of local and temporal adaptation in large lake *Daphnia* species" ***M. Herrmann**, P. Kreuzer, M. Cordellier, N. Henning, K. Schwenk*, 7th symposium on Eco-evolutionary Dynamics 2013, Leuven, Belgium

„Grubbing in mud – has climate change already left traces in *Daphnias* genome?" *P. Kreuzer, **M. Herrmann**, M. Cordellier, B. Streit, N. Brede, K. Schwenk,* The Daphnia Genomics Consortium Meeting 2010, Leuven, Belgium

Scholarships

since February 2012	Senckenberg Stipend for a PhD fellowship at the Goethe University in Frankfurt am Main and the Biodiversity and Climate Research Centre (Bik-F) in Frankfurt am Main, granted by the Senckenberg Gesellschaft für Naturforschung
September 2012	Travel grant by the Hermann-Willkomm Foundation Frankfurt am Main

Declaration

I herewith declare that this thesis reports my independent work on "Evolutionary Genomics in *Daphnia*". I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included and that I made no use of a consulting service or any other form of paid assistance. All used resources are specified and contributions from other authors and third parties are fully acknowledged.

This or another thesis has never been previously submitted in part or in whole to another academic or non-academic institution as an examination work or in any other form.

I am aware of and understand that violations of the above mentioned items may result in a revocation of the doctoral degree as well as in legal consequences.

Place, date

Signature

Acknowledgements

First and foremost, I would like to thank my supervisor Klaus Schwenk for his support and guidance, but also for his enthusiasm, optimism, and patience with me through all the years. The word always looked brighter after discussing results with you!
I am also grateful to my second supervisor Mathilde Cordellier for her support and valuable input along the way, but also for her healthy skepticism and brining the Nähkränzchen to life.
Thanks to both of you for your friendship. I feel very lucky for our paths crossing and for you having accompanied me on this journey.

Many thanks go to my colleagues and co-authors of my thesis chapters, Nicole Henning-Lucass, Pia Kreuzer and Suda Parimala Ravindran. It was a great time and I wish we could have worked together longer. Nicole and Pia, thank you for the fond memories of our diploma thesis and PhD study times and especially our crazy sampling trip(s). I'm always ready to get stuck on a lake with you again.

Very special thanks go to Anne Thielsch, who introduced me into the world of population genetics, and to Verena Kutschera, who ventured into the world of genomics with me. Thank you both for your friendship, advice, encouragement and inspiring discussions! You are the best travel companions in the world and your input during writing helped me so much. My time as a PhD student would have been much less fun without both you. Anne, I admire that you always know where you stand. Thank you and Martin so much for welcoming me to your home in Landau, you two are the most generous people and simply the best. Verena, I really miss you! Thank you for being something like my working group substitute for me in Bik-F. You are my favorite person to discuss –omics with (amongst many other things of course) and I'm just really glad we met and stay in contact.

In my time in the Siesmayerstraße, in Bik-F and in Landau, I met many great colleagues and friends for life. Thank you Elke, Ruth, Nicole, Simit, Barbara, Frank, all the Susannes, Melanie, Julia and all the others I met along the way, with whom I had the pleasure to work and discuss ideas, but also to drink cocktails, cook, sew, and have fun!

Ganz besonders möchte ich meinen Eltern für all ihre Liebe und Unterstützung danken. Danke, dass ihr mehr Geduld mit mir habt als ich, jeden kleinen Fortschritt mit mir gefeiert habt und für das viele gute Zureden bei anregenden Getränken. Danke, dass ihr immer für mich da seid und dass ihr mir alles zutraut.

And last but not least, Can, thank you for picking me up every time I was stressed out, for encouraging me endlessly, but also for being my very own and favorite personal IT support! Thank you for everything. You are the light of my life.