Received: 29 September 2018 | Revised: 5 December 2018 | Accepted: 13 December 2018

DOI: 10.1002/ece3.4894

ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

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

Suda Parimala Ravindran[1] | Maike Herrmann[2] | Mathilde Cordellier[1]

[1]Institute of Zoology, Universität Hamburg, Hamburg, Germany

[2]Department of Veterinary Medicine, Paul-Ehrlich-Institut, Langen, Germany

Correspondence
Mathilde Cordellier, Institute of Zoology, Universität Hamburg, Hamburg, Germany.
Email: mathilde.cordellier@uni-hamburg.de

Funding information
Volkswagen Stiftung, Grant/Award Number: 86030

Abstract

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

KEYWORDS

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

1 | INTRODUCTION

Natural genetic variation shapes divergence in phenotypic traits and is an important resource for evolutionary processes (Oleksiak, Churchill, & Crawford, 2002). Populations respond to environmental variation by genetically adapting to their environments (Hereford, 2009; Kawecki & Ebert, 2004; Savolainen, Lascoux, & Merila, 2013), often showing variations at both gene expression and sequence level across the geographic range of a species. One of the fundamental goals of research in the field of molecular evolution is to resolve the evolutionary processes driving the rise and maintenance of expression and sequence polymorphisms behind this variation. Revealing their effect on an organism's fitness thereby aids to understand the genetic basis of local adaptation (MacManes & Eisen, 2014). Gene expression patterns link genotypes and phenotypes, sometimes called a "molecular phenotype," and as such is an important component in local adaptation processes (Lopez-Maury, Marguerat, & Bahler, 2008). Several studies have reported the testing of different

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

populations exposed to different treatments and examining their transcriptional response, for example, in springtails (*Folsomia* [De Boer et al., 2013] and *Orchesella* [Roelofs et al., 2009]), oyster (*Crassostrea virginica*; Chapman et al., 2011; Chapman et al., 2009), sparrows (*Zonotrichia capensis*; Cheviron, Whitehead, & Brumfield, 2008), flounder (*Platichthys flesus*; Larsen, Nielsen, Williams, & Loeschcke, 2008), and seagrass (*Zostera marina*; Jueterbock et al., 2016; Reusch et al., 2008), thereby identifying candidate genes involved in local adaptation. Gene expression variation can be highly heritable (Brem & Kruglyak, 2005; Schadt et al., 2003; Whitehead & Crawford, 2006b). Moreover, constitutive gene expression patterns also differ within- and among-natural populations (e.g., Roberge, Guderley, & Bernatchez, 2007; Whitehead & Crawford, 2006a), strongly suggesting that standing variation in constitutive gene expression is shaped by local adaptation. Natural selection acts immediately on newly arisen variation (in contrast to adaptation observed from standing genetic variation) as there are neutral and slightly deleterious variations preserved in a population, which may become beneficial upon changes in selection regimes (Barrett & Schluter, 2008). After a sudden change of environment, standing variation can contribute to fast adaptation (Feulner et al., 2013; Kitano et al., 2008). Identifying allelic/genetic variants underlying differences in expression profiles can be helpful in hypothesizing gene functions (Jansen & Nap, 2001; Kesari et al., 2012; Rockman, 2008). Although prior knowledge of the specific loci is not a prerequisite to learn about adaptive processes in most cases, identification of genetic features underlying local adaptation is critical in answering fundamental questions about natural selection (Rausher & Delph, 2015).

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

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

FIGURE 1 Waterfleas, *Daphnia galeata*, Photograph: Mathilde Cordellier

Genetic variation has been reported for numerous traits in *Daphnia*, such as life history traits (e.g., Henning-Lucass, Cordellier, Streit, & Schwenk, 2016), vertical migration (e.g., Haupt, Stockenreiter, Baumgartner, Boersma, & Stibor, 2009), fish escape behavior (e.g., Pietrzak, Pijanowska, & Dawidowicz, 2015), resistance against parasites (e.g., Routtu & Ebert, 2015), and immune response (e.g., Garbutt, O'Donoghue, McTaggart, Wilson, & Little, 2014). Furthermore, it was shown that responses to many chemical stressors such as phosphorus (Roy Chowdhury et al., 2015; Roy Chowdhury, Lopez, Weider, Colbourne, & Jeyasingh, 2014), copper (Poynton, Zuzow, Loguinov, Perkins, & Vulpe, 2008), cadmium (Soetaert et al., 2007), and pharmaceutical products like ibuprofen (Hayashi, Heckmann, Callaghan, & Sibly, 2008; Heckmann, Callaghan, & Hooper, 2007) have a genetic basis as well. Within- and between-population comparisons in *Daphnia* have been

conducted extensively using varied environmental perturbations providing evidence for local adaptation (e.g., Barata, Baird, Mitchell, & Soares, 2002; Declerck, Cousyn, & De Meester, 2001; Ebert, Zschokke-Rohringer, & Carius, 1998; Spitze, 1993). Although various aspects like phylogeography, functional morphology, physiology, and life history evolution have been in the limelight of *Daphnia* research for several decades (Eads et al., 2008), *Daphnia* genomics investigations have begun only in the last decade with the availability of the *Daphnia pulex* genome (Colbourne et al., 2011). A considerable number of studies (e.g., Bento et al., 2017; Miner, De Meester, Pfrender, Lampert, & Hairston, 2012; Orsini et al., 2016; Yampolsky et al., 2014) on biotic and abiotic factors have been carried out showing how *Daphnia* respond to environmental perturbations by changes in gene expression. However, little is known about the intraspecific variability at the gene expression level in *Daphnia*, since the above-mentioned studies focused on stressor-driven responses using a limited number of genotypes.

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

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

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

2 | METHODS

2.1 | Sampling and RNA collection

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

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

2.2 | RNA preparation and sequencing

Total RNA was extracted following a modified phenol/chloroform protocol and further processed using the RNeasy kit (Qiagen, Hilden, Germany). The total RNA was eluted in RNAse-free water, and the concentration and quality (RNA integrity number and phenol contamination) were checked using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The 72 total RNA samples (4 lakes × 6 genotypes × 3 biological replicates) were sent to the company GATC (Konstanz, Germany) for library preparation and sequencing. Following reverse transcription and cDNA construction using random primers, 50 bp single-end (SE) reads were sequenced on an Illumina HiSeq 2000 (San Diego, CA, USA). To avoid block effects and confounding effects in the downstream analysis, we used a completely randomized design; each library was sequenced on at least two different lanes, on a total of nine lanes. Detailed information can be found in Supporting Information Table S1.

2.3 | Quality trimming, mapping, and read counts

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

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

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

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

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

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

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

Using microsatellite allelic frequencies, the coancestry coefficients by admixture F model were calculated using "do.all" function implemented in the RAFM package (Karhunen & Ovaskainen, 2012). We ran a total of 200,000 iterations with thinning at an interval of 1,000 and discarded the first 1,000 iterations as "burn-in." The output was a list which contained samples from the posterior distributions of allele frequencies. Values from the posterior coancestry matrix, "theta," were used as input for the Metropolis-Hastings (MH) algorithm along with the normalized "mean" read counts of the replicates (i.e., one value per genotype) for each DET as implemented in DRIFTSEL. We ran a total of 5,000 iterations with thinning at 1,000 samples and discarded the first 100 iterations as burn-in. The output of MH algorithm was a matrix of posterior of subpopulation effects (pop.ef), used to estimate the H.test values. The H.test describes whether the population means correlate with the genetic data more than it would be expected on basis of shared evolutionary history. Large *H*-values (*H*-value ≥ 0.95) imply that the populations are more locally adapted than expected by chance.

2.5 | Intra- and interpopulation variation

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

2.6 | Variant calling and filtering

The variant calling and filtering steps have already been described in Herrmann, Ravindran, Schwenk, and Cordellier (2017b). Briefly, the aligned reads from RNA-seq data were merged using samtools (Li et al., 2009). GATK (McKenna et al., 2010) was used to split exon segments and reassign the mapping qualities (SplitNCigarReads), and indels were aligned (RealignerTargetCreator and IndelRealigner). The HaplotypeCaller (DePristo et al., 2011) function was used for the initial variant calls for the realigned reads, and samples were jointly genotyped using GATK's GenotypeGVCFs tool. A single vcf file was created, and false-positive variant calls were filtered with the following criteria: (a) clusterWindowSize = 35; (b) Quality by depth

(QD) < 2.0; (c) Fisher Strand (FS) > 30.0. This produced a variant dataset with not only biallelic variants but also triallelic variants and indels.

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

2.7 | Neutrality statistics

To obtain alignments of transcript sequences, SNP calling datasets were filtered as described above. Beagle 4.1 (Browning & Browning, 2007) was used to phase SNP calling data, and a python script (available upon request) was used to parse the phased vcf file to sample sequences in fasta format. After phasing, we obtained 13,006 transcripts containing SNPs and the sequences were input in R. A multiple sequence alignment and Tajima's *D* statistics (with *p*-values) were obtained population-wise for each transcript using the pegas package (Paradis, 2010) in R.

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

2.8 | Inbreeding coefficient and mutation frequencies

The inbreeding coefficient for the final SNP dataset was calculated with the –het function in VCFtools (Danecek et al., 2011). The ratio between the expected heterozygosity (H_E) and observed heterozygosity (H_O) was calculated based on available SNP information, and plots were created using ggplot2 (Wickham, 2016) in R.

2.9 | Sequence versus regulatory variation

To visualize the proportion of transcripts responsible for local adaptation at regulatory and sequence level, we consolidated the list of transcripts from various analyses as performed above and represented it with an alluvial diagram (http://rawgraphs.io/). In an alluvial diagram, each black rectangle is called a "node," the colored regions linking the nodes are called "flows," and the vertical group of nodes is called "steps." In our analyses, we had four steps: DESeq2, DRIFTSEL, LOSITAN, and Tajima's *D* (Figure 4).

2.10 | Annotation and functional analysis

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

We classified the orthoMCL clusters into the following categories:

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

2.11 | Inparalogs and misassemblies

To assess whether *D. galeata* DETs in an orthologous group are "inparalogs," isoforms or the result of misassembly, we computed the pairwise sequence divergence for those orthoMCL clusters containing DETs from at least two different populations. Since each significantly differentially expressed transcript was assigned as a DET only to the population in which it was upregulated the most, clusters containing more than one DET most likely contained DETs from different populations. Based on the number of populations within their orthoMCL cluster, the DETs were classified into the categories: "1Pop," "2Pop," "3Pop," and "4Pop," and unclustered DETs were categorized as "no-cluster DETs." "No-cluster DETs" and "1Pop" DETs were excluded from further analysis. In total, there were 716 orthoMCL clusters that contained DETs from at least two populations. Pairwise alignments of the amino acid sequences in each orthologous group were performed using the iterative refinement method incorporating local pairwise alignment information (L-INS-i)

in MAFFT (Katoh, Misawa, Kuma, & Miyata, 2002). We then used EMBOSS tranalign (Rice, Longden, & Bleasby, 2000) to generate alignments of nucleic acid coding regions translated from aligned protein sequences. Pairwise genetic divergence was computed with "dist.dna" function implemented in the ape package (Paradis, Claude, & Strimmer, 2004) in R, using the Kimura-2-parameters model with gamma correction. We used an arbitrary cutoff value of 2 to distinguish inparalogs from misassembled sequences.

2.12 | Gene ontology (GO) enrichment analysis

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

2.13 | Weighted gene co-expression network analysis

To gain insights into the population-specific regulatory patterns of transcripts in *D. galeata*, we performed a weighted gene co-expression network analysis with WGCNA (Langfelder & Horvath, 2008) using the variance stabilized normalized read counts obtained in DESeq2 analysis. Transcripts and samples that had lower expression values were excluded from every population using the "goodSampleGenes" function in WGCNA and used for downstream analysis. In total, 32,375 transcripts were used for the construction of gene co-expression networks. To identify population-specific co-expression modules (i.e., clusters of highly correlated transcripts), a network was first built using the full dataset (i.e., with samples and transcripts from all populations) and one network for each population using expression values specific to all genotype and biological replicates. The population-specific network was compared to the reference network, and an adjacency matrix was calculated. Clusters were identified using the WGCNA Topological Overlap Matrices (TOM). For every transcript and module detected automatically, WGCNA assigns a color based on the module membership (MM) value. An MM value is a measure of module membership which is obtained by correlating its gene expression profiles with module eigengene (i.e., the first principal component of a given module). For example, if a transcript has an MMred value close to ±1, the transcript is assigned to the red module (Langfelder & Horvath, 2008). Each module is assigned a color based on the module size: "turquoise" denotes the largest module, blue next, followed by brown, green, yellow, and so on. The color "grey" is reserved for unassigned transcripts (Langfelder & Horvath, 2008). Similarly, the module "gold" consists of 1,000 randomly selected transcripts that represent a sample of the whole network and statistical measures have no meaning for this module (Langfelder & Horvath, 2008).

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

3 | RESULTS

3.1 | Sequencing results and mapping statistics

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

3.2 | Differential expression

The intraspecific variation in transcript expression in the four populations was visualized from a read counts matrix of the 32,903 transcripts using PCA (Figure 2a). A large proportion of the observed variance (19%) is explained by the first principal component (PC1). PC2 and PC3 explained 12% and 10% of the total variance, respectively. Clear population clustering is evident along PC2 for Pop.M and in Pop.J except for two genotypes (J2.1 and J2.4). However, genotypes from Pop.G and Pop.LC belong to overlapping clusters (Figure 2a and Supporting Information Appendix S1). No evident clustering according to experimental parameters (i.e., culture conditions, harvesting, RNA extraction batches) was visible on the PCA.

After conducting pairwise contrast analyses with DESeq2, we identified transcripts exclusively upregulated for each population when compared to all others ($p_{adj} \leq 0.05$; thereafter DETs). In total, 10,820 of 32,903 transcripts (~33%) showed significant expression differences in pairwise comparisons between populations. Of all ~33,000 transcripts, 9.6%, 8.1%, 7.2%, and 7.8% were population-specific DETs for the populations Pop.G, Pop.J, Pop.LC, and Pop.M, respectively (Figure 2b).

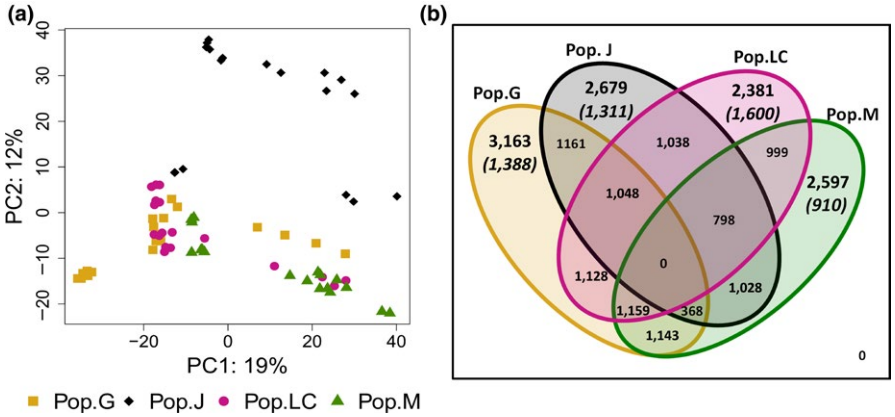

FIGURE 2 Gene expression patterns. (a) Gene expression PCA of the four sampled populations: Pop.G (Lake Greifensee), Pop.J (Jordan Reservoir), Pop.LC (Lake Constance), and Pop.M (Müggelsee). Percentages on the *x*- and *y*-axis indicate the percentage of variance explained by each principal component. (b) Venn diagram illustrating the number of differentially expressed transcripts (DET) between the four populations. Numbers in brackets indicate the number of transcripts deviating from the neutral expectations according to the DRIFTSEL analysis

3.3 | Role of natural selection on transcript expression levels

The DRIFTSEL multivariate approach was used to identify transcripts for which the observed differential expression could not be explained by phylogenetic distance and genetic drift alone; the alternative explanation being that the observed divergence would be attributable to selection and therefore possibly to local adaptation events. In total, 48% of 10,820 differentially expressed transcripts showed greater differential expression than expected under neutrality (H-value ≥0.95, Figure 2b and Supporting Information Table S2), indicating that the observed pattern is due to local adaptation for these transcripts. Pop.LC had the highest number of DETs deviating from the neutral expectations (67% transcripts out of 2,381), followed by Pop.G (~44% out of 3,163), Pop.J (~49% out of 2,679), and Pop.M (35% out of 2,597).

3.4 | Expression variation among individuals and populations

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

3.5 | Sequence-based divergence

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

A PCA was carried out based on a matrix of all biallelic SNP sites to illustrate the population structure among the four populations. Although PC1 explained the maximum variance (12%) (Figure 3a) and four distinct clusters corresponding to the populations were seen against PC2. PC2 and PC3 each explained 8% of the variance (Supporting Information Appendix S2). PC2 clearly separated the genotypes belonging to Pop.J from the remainder of the data (Figure 3a).

3.6 | Inbreeding coefficient

The inbreeding coefficient values ranged from −1.19 to 0.22 for genotypes from Pop.G, from 0.08 to 0.17 for those from Pop.J, from 0.07 to 0.15 within Pop.LC, and from −0.06 to 0.32 for those belonging to Pop.M (Figure 3b). Within Pop.G, four out of the six genotypes were more heterozygous than expected, indicating outbreeding. In Pop.J, Pop.LC, and Pop.M (except genotype M9 therein), there were less heterozygous than expected, implying inbreeding.

3.7 | Sequence evolution

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

TABLE 1 Summary of SNP data

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

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

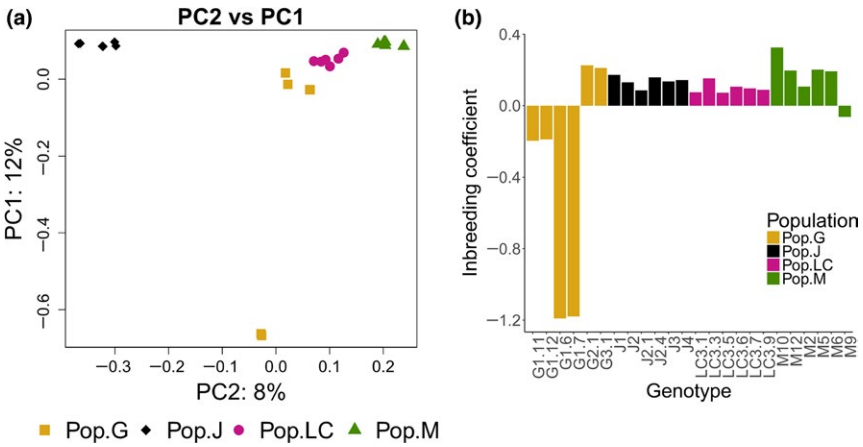

FIGURE 3 SNP patterns and inbreeding coefficient. (a) SNP PCA of the four sampled populations: Pop.G (Lake Greifensee), Pop.J (Jordan reservoir), Pop.LC (Lake Constance), and Pop.M (Müggelsee). Percentages on the *x*- and *y*-axis indicate the percentage of variance explained by each principal component. (b) Barplot illustrating the inbreeding coefficient for each genotype

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

3.8 | Sequence versus regulatory variation

The proportion of transcripts identified to be candidates for local adaptation at both sequence and regulatory level were visualized using a flow diagram (Figure 4). Among the 10,820 transcripts identified to be differentially expressed, ~46% showed signs of selection at the regulatory level according to DRIFTSEL. Of these, ~15% were identified as outliers under balancing and/or diversifying selection in LOSITAN. About 26% of these outliers had a significantly negative or positive Tajima's *D* value in at least one population, which might be attributed to selection but can also stem from other evolutionary processes such as population growth, reduction or subdivision, bottleneck events, and migration.

TABLE 2 Tajima's *D* test for selection

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

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

3.9 | Functional annotation

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

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

For identifying *Daphnia*-specific orthologs and those that share orthology with other arthropods, the orthoMCL data were classified into six categories (as described in the Methods section); 3,058 orthology clusters (of which 1,735 clusters contained DETs) were containing exclusively *D. galeata* transcripts, 985 clusters (of which 543 clusters contained DETs) contained only *D. galeata* transcripts and *D. pulex* genes, and 651 clusters (including 224 DETs) contained only *D. galeata* and *D. magna* transcripts; 3,336 orthoMCL clusters (of which 1,239 clusters contained *D. galeata* DETs) contained all three *Daphnia* species used in the analysis. Furthermore, 12 clusters (four clusters containing DETs) were containing *D. galeata* transcripts along with two other arthropods (*D. melanogaster* and *N. vitripennis*). In total, 4,657 clusters (1,586 clusters containing DETs) contained transcripts/genes for all five species (three *Daphnia* species and two insects) used in the present study (Supporting Information Appendix S3c, Table S5).

3.10 | Assessment of assembly artifacts and inparalogs

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

3.11 | Gene ontology enrichment analysis

GO enrichment analysis was performed on the candidate transcripts as identified from DRIFTSEL (*H* value ≥0.95) and Tajima's *D* analyses. We observed an enrichment for several metabolic processes such as ATP binding, DNA binding, microtubule binding, transporter activities, and signaling pathways (Supporting Information Table S6) in both analyses in all population-specific sets. Specifically, in Pop.G, DRIFTSEL and Tajima's *D* analysis had five GO terms in common; in

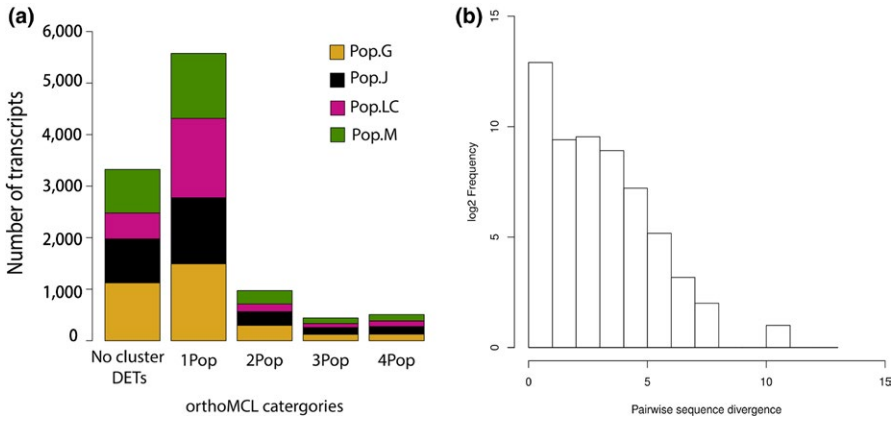

FIGURE 4 Differentiating misassembly from inparalogs. (a) Barplot showing the number of DETs co-occurring with DETs from other populations within an orthoMCL cluster. "No-cluster DETs" refers to DETs not assigned to an orthoMCL cluster. 1Pop, 2Pop, 3Pop, and 4Pop refer to DETs found in orthoMCL clusters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

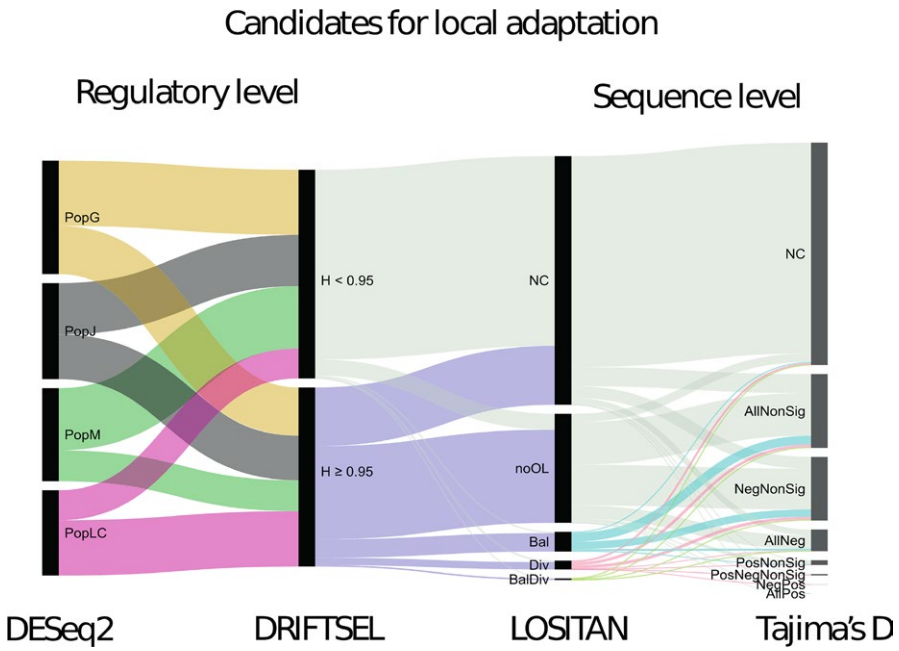

FIGURE 5 Flow diagram representing the proportion of transcripts that are candidates for local adaptation at the regulatory and sequence level. Each analysis or "step" is represented by a vertical group of black rectangle bars, called nodes. The colored areas linking the nodes are called "flows." The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

Pop.J, they had one GO term in common; in Pop.LC, they had four GO terms in common; and in Pop.M, they had seven GO terms in common.

3.12 | Weighted gene co-expression network analysis

The WGCNA on 32,375 transcripts identified 29 co-expression modules (Figure 6) in the reference network (see Section 2). We observed varying numbers of modules and transcripts clustered in each population-specific network (Supporting Information Table S7a–d). However, after assessing the conserved modules, where each population-specific network was compared to the reference network, 24 modules (out of 29) were well conserved (Z-score ≥10) among the populations. The conserved modules included 10,256 transcripts altogether, which is about 31% of all transcripts in *D. galeata*, with the largest module, "turquoise" including 2,857 transcripts. Two modules (grey and gold) with uncharacterized and random transcripts contained 16,600 and 1,000 transcripts, respectively. These results are consistent with the gene expression analysis that showed little differences between the populations.

4 | DISCUSSION

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

4.1 | Population divergence at the sequence level

SNPs became the absolute marker of choice for molecular genetic analysis as the mining of polymorphisms is the cheapest source for

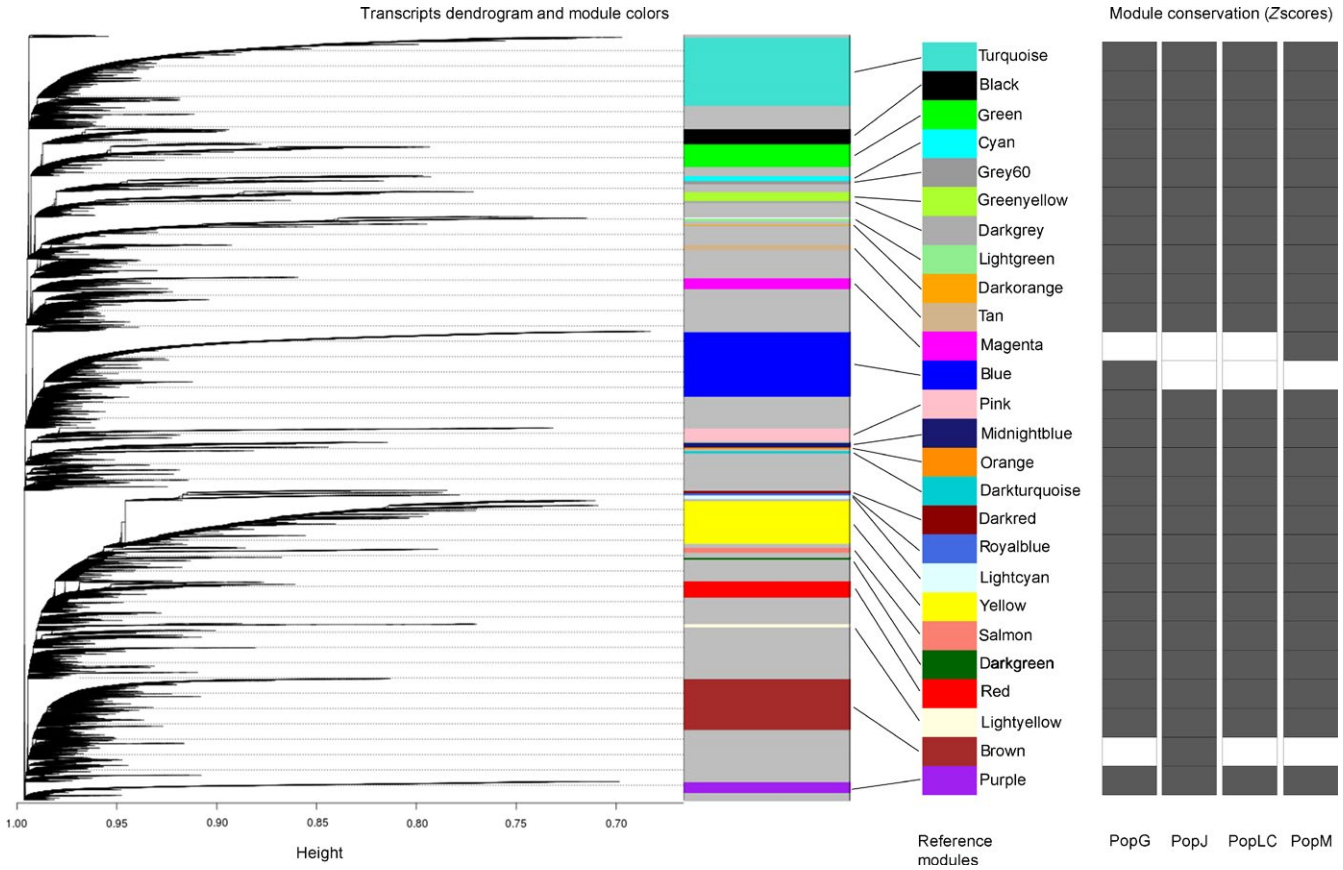

FIGURE 6 Cluster dendrogram of transcripts for the reference network in *Daphnia galeata*, with dissimilarity based on the topological overlap matrix (TOM). The co-expression modules are colored in an arbitrary way by the WGCNA package, and the size of the bar is proportional to the number of transcripts in the module. The right-hand side grid represents the module conservation in each population. Modules with a *Z*-score ≤10 are shown in white, and modules with a *Z*-score ≥10 are colored in dark grey

genetic variability (Taillon Miller, Gu, Li, Hillier, & Kwok, 1998). Our PCA analysis on SNP data revealed four clear population clusters, and our results are in agreement with a highly structured population model across the transcriptome. Although two of the genotypes (G1.6 and G1.7) from Pop.G were located outside the Pop.G cluster in the PCA plot, the populations were clearly distinguished and corresponded to the four lakes sampled. This pattern might be the result of several nonexclusive phenomena: initial founder effects, isolation-by-distance and genetic drift, and natural divergent selection, since the studied populations originate from lakes located in different ecoregions.

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

All the phenomena cited above have an impact on population structure across the genome and might mask highly diverging loci resulting from natural selection. We assessed patterns of divergence at the sequence level through neutrality tests (Tajima's *D*). This suggested that all populations of *Daphnia* examined in this study had a substantial amount (~48% transcripts) of loci with an excess of low-frequency polymorphisms (i.e., $D < 0$) relative to the neutral expectation. This pattern may result from positive selection, a bottleneck, or population expansion. It is consistent with previous observations in *Daphnia* from Lake Greifensee and Lake Constance (Brede et al., 2009) and crustacean zooplankton from Lake Constance (Straile & Walter, 1998) which have all undergone historical bottleneck events. Similarly, Lake Müggelsee, a large shallow lake, has undergone severe bottlenecks due to increased turbidity and because vegetation disappeared almost completely after the 1960s (Okun, Mendonca, & Mehner, 2005). One other explanation for the excess of rare alleles is selection against genotypes carrying deleterious alleles.

Although a high frequency of rare polymorphisms was observed in our analysis, there were few transcripts (~1.7% transcripts) that had a lower frequency of rare alleles ($D > 0$) in the four populations, indicating that some loci are either under balancing selection (where heterozygous genotypes are favored) or under diversifying selection (where genotypes carrying the less common alleles are favored). A lower frequency of rare alleles also occurs if there is a recent population admixture (Stajich & Hahn, 2005). This argument is consistent with our inbreeding coefficient measures. Most of the genotypes in population G, as well as M9, were more heterozygous than expected. While genotype M9 from Müggelsee might be an exception, the pattern observed in Greifensee could be the consequence of outbreeding and/or high genetic variability in this population. This might also stem from past hybridization events (Brede et al., 2009). Although our primary checks for this with a handful of microsatellite markers did not flag the genotypes as having a hybrid origin, the use of high-density markers such as the SNPs might have uncovered traces of introgression. Under the Hardy–Weinberg equilibrium, genotypes G2.1 and G3.1 from Pop.G, all genotypes in Pop.J and Pop.LC, and all genotypes except M9 in Pop.M show that the observed heterozygosity was less than the expected heterozygosity, which is an indication of lower genetic variability and inbreeding. Such low heterozygosity patterns at the individual level can be attributed to inbreeding (Keller & Donald, 2002), but also due to a lack of variation in the source population, either caused by a small founder population size or a severe bottleneck during population history (Luikart, Allendorf, Cornuet, & Sherwin, 1998). Further, the ecology and growth dynamics of *Daphnia* populations might exacerbate the founder effects. After an initial hatching phase from the resting eggs bank and exponential population growth in the spring, clonal selection occurs throughout the growing season (Vanoverbeke & De Meester, 1997). Therefore, it is possible that only a few clonal lines contribute to the resting eggs population each year. However, while a reduced number of clonal lines might contribute to the yearly "archiving" of genetic diversity, two processes counteract the immediate diversity loss. First, the spring recruitment does not only rely on eggs from the previous year but rather on a mixture (Vanoverbeke & De Meester, 1997), and might even integrate overwintering clones in larger permanent lakes (but see Yin, Gießler, Griebel, & Wolinska, 2014 for an overview). Second, clonal erosion does not affect the same genotypes every year, leading to year-to-year heterogeneity, such as the one observed in the long-term study by Griebel, Giessler, Yin, and Wolinska (2016). Clonal erosion thus does not necessarily lead to a downward spiral of genetic diversity loss, and the high stochasticity of both clonal selection and hatching ensures a preservation of the genetic diversity in every habitat.

4.2 | Gene expression variability and signals of selection

While the patterns observed at the sequence level tend to support the role of genetic drift, founder, and monopolization effects in shaping the observed patterns, the results of our gene expression analysis delivered a mixed message. This was evident in the PCA based on the gene expression data, where no distinct clusters corresponding to populations are clearly visible. This observation was consistent with our network co-expression analysis which showed that the identified modules are conserved in all populations (Figure 6), with a few exceptions. The analysis of variance confirms this finding, with a relatively low number of transcripts for which the mean read counts differ significantly between populations and not between genotypes. While studies on differential expression in *Gliricidia sepium* (Chalmers, Waugh, Sprent, & Simons, 1992) and *Arabidopsis halleri* (Macnair, 2002) have observed substantial between population variance at the gene expression level, our results are consistent with several studies, for example, on fish (*Fundulus heteroclitus*; Whitehead & Crawford, 2006a) and snails (*Melanoides tuberculata*; Facon, Pointier, Jarne, Sarda, & David, 2008) which showed large within-population variation. Additionally, numerous studies on life history traits in *Daphnia* also report very high intrapopulation variability (Beckerman, Rodgers, & Dennis, 2010; Castro, Consciência, & Gonçalves, 2007; Cousyn et al., 2001; Macháček, 1991). A common garden experiment conducted on the very same clonal lines also showed a higher phenotypic variability within populations than among populations (Tams, Luneburg, Seddar, Detampel, & Cordellier, 2018). Finally, the observed relative homogeneity in the gene expression profiles might be the consequence of high selective pressure on transcription regulation or canalization (Waddington, 1942). Such canalization allows for storage of cryptic genetic variation that would be uncovered in stress response assessments. However, our experimental setup was designed to avoid stress, and transcriptome characterization of the same genotypes under conditions mimicking predation, parasite, or food stress, for example, might reveal a greater divergence between the populations.

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

4.3 | Sequence versus regulatory variation in *Daphnia galeata*

Correlating expression profiles with sequence divergence helps to identify transcripts that are potentially under the influence of local adaptation at both gene expression and sequence level. Linking gene expression profiles with sequence polymorphisms and their associated functions aids in understanding the genetic basis of adaptation as seen in the desert-adapted mouse (*Peromyscus eremicus*; MacManes & Eisen, 2014) and in the Patagonian olive mouse (*Abrothrix olivacea*; Giorello et al., 2018). Our results revealed ~30% of the transcripts to share divergence at both sequence and regulatory level (Figure 4). There are two possible explanations for the observed differences in sequence and regulatory level variation (Hodgins, Yeaman, Nurkowski, Rieseberg, & Aitken, 2016). The first is that there is an increase in the rate of fixation due to transcripts under positive selection and divergence in expression patterns. For example, variation in gene expression might lead to selection for sequence variation to improve the functional role of the transcript in its altered role (Hodgins et al., 2016). A second explanation is that the differentially expressed transcripts may experience reduced negative selection in one or all four populations. For instance, higher transcript expression is associated with greater negative selection. Hence, a reduction in transcript expression in one population compared to others may be accompanied by relaxation of selection in that population.

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

4.4 | Gene annotation and evaluation of inparalogs

Gene annotation is quite challenging in organisms lacking reference genomes, and functional annotation then relies on the availability of transcriptomic sequences from the closest available taxon. In this study, we were able to annotate 66.5% of the transcripts using BLAST analysis (Supporting Information Appendix S3a). However, many of the transcripts were homologous to a *D. pulex* "hypothetical protein," likely because (a) they are similar in function to noncoding regions or pseudogenes or (b) novel coding transcripts that are yet to be functionally characterized (Vatanparast et al., 2016). Furthermore, we were able to predict domains for 80% of the transcripts using Pfam analysis (Supporting Information Appendix S3b, Table S4). Our orthoMCL results (Supporting Information Appendix S3c, Table S4) showed that several (~45%) of the *D. galeata* transcripts were orthologous to one or all species of *Daphnia* used for comparison, indicating that the genes/transcripts have all evolved from a common *Daphnia*-specific ancestral gene via speciation. In addition to this, ~25% of *Daphnia* genes/transcripts are orthologous to two insect species (*D. melanogaster* and *N. vitripennis*). Our level of unannotated transcripts is similar to results reported from other organisms lacking extensive genomic resources, for example, from plants like field pea (*Pisum sativum;* Sudheesh et al., 2015), chickpea (*Cicer arietinum;* Kudapa et al., 2014), and winged bean (*Psophocarpus tetragonolobus*; Vatanparast et al., 2016). This limited our interpretation of the functional role of *Daphnia* transcripts and thereby their associations to known ecological stressors. A second issue raised when lacking a reference genome is that it might be difficult to tease apart inparalogs created by duplication events, isoforms, and even misassemblies, leading to an artificially inflated number of similar sequences for each distinct gene in the transcript set. Only ~18% of the population-specific DETs had one or more putative paralogs also identified as differentially expressed in at least one other population. For DETs from two or more populations that co-occurred in orthoMCL clusters, we were able to distinguish between actual paralogs (transcript pairs that had a sequence divergence value >2, Figure 5b) and transcripts with sequence divergence value <2. Genomic information is now required for this species in order to accurately assign transcripts to genes and correctly assess whether two different populations might indeed express different gene copies with similar functions.

5 | FUTURE DIRECTIONS AND CONCLUSIONS

In summary, we described here an approach that combines both transcriptomic expression profiles and sequence information to understand local adaptation in *D. galeata*. Although the set of transcripts contributing to population divergence at the sequence and the expression level differs, both levels constitute alternative routes for responding to selection pressures (Pai, Pritchard, & Gilad, 2015), showing that these transcripts can contribute to local adaptation and paving way for future research. From our functional analysis, it was evident that most of our transcripts were *Daphnia* specific although they had hypothetical functions. To understand the function of the hypothetical transcripts in *D. galeata* and their response to environmental perturbations, a comparative approach using the gene expression data from numerous other *Daphnia* studies should be used. Although we noticed correlations between expression patterns and sequence divergence for the *D. galeata* transcripts, we lack genomic and phylogenetic information. This information may help "bridge the gap" for understanding the relative roles of positive or

negative selection in driving coding sequence and gene expression divergence.

ACKNOWLEDGMENTS

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

CONFLICT OF INTEREST

None declared.

AUTHOR'S CONTRIBUTIONS

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

DATA ACCESSIBILITY

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

The raw read counts used as input for differential transcript expression, results for the pairwise contrast analysis conducted in DESEq2, and the number of variants per sample before and after filtering, number of variant sites per transcript are all available on DRYAD in Supporting Information Tables S10, S11, and S3–S4, respectively (https://datadryad.org//resource/doi:10.5061/dryad.p85m5).

The VCF file is now available on Dryad Digital Repository: https://doi.org/10.5061/dryad.56jt85j.

ORCID

Suda Parimala Ravindran https://orcid.org/0000-0002-6745-4977

Maike Herrmann https://orcid.org/0000-0002-0832-3539

Mathilde Cordellier https://orcid.org/0000-0001-7376-4560

REFERENCES

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

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics*, *9*, 323. https://doi.org/10.1186/1471-2105-9-323

Barata, C., Baird, D. J., Mitchell, S. E., & Soares, A. M. V. M. (2002). Among- and within-population variability in tolerance to cadmium stress in natural populations of *Daphnia magna*. *Environmental Toxicology and Chemistry*, *21*, 1058–1064.

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

Beckerman, A. P., Rodgers, G. M., & Dennis, S. R. (2010). The reaction norm of size and age at maturity under multiple predator risk. *Journal of Animal Ecology*, *79*, 1069–1076. https://doi.org/10.1111/j.1365-2656.2010.01703.x

Benjamini, Y., & Yosef, H. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, *57*, 289–300.

Bento, G., Routtu, J., Fields, P. D., Bourgeois, Y., Du Pasquier, L., & Ebert, D. (2017). The genetic basis of resistance and matching-allele interactions of a host-parasite system: The *Daphnia magna-Pasteuria ramosa* model. *PLoS Genetics*, *13*, e1006596. https://doi.org/10.1371/journal.pgen.1006596

Boileau, M. G., Hebert, P. D. N., & Schwartz, S. S. (1992). Non-equilibrium gene frequency divergence: Persistent founder effects in natural populations. *Journal of Evolutionary Biology*, *5*, 25–39. https://doi.org/10.1046/j.1420-9101.1992.5010025.x

Böndel, K. B., Lainer, H., Nosenko, T., Mboup, M., Tellier, A., & Stephan, W. (2015). North-South colonization associated with local adaptation of the wild tomato species *Solanum chilense*. *Molecular Biology and Evolution*, *32*, 2932–2943.

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

Brem, R. B., & Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 1572–1577. https://doi.org/10.1073/pnas.0408709102

Browning, S. R., & Browning, B. L. (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. https://doi.org/10.1086/521987

Castro, B. B., Consciência, S., & Gonçalves, F. (2007). Life history responses of *Daphnia longispina* to mosquitofish (*Gambusia holbrooki*) and pumpkinseed (*Lepomis gibbosus*) kairomones. *Hydrobiologia*, *594*, 165–174. https://doi.org/10.1007/s10750-007-9074-5

Chalmers, K. J., Waugh, R., Sprent, J. I., Simons, A. J., & Powell, W. (1992). Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. *Heredity*, *69*, 465–472.

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., … Didonato, G. (2009). A transcriptomic analysis of land-use impacts on the oyster, *Crassostrea virginica*, in the South Atlantic bight. *Molecular Ecology*, *18*, 2415–2425.

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., … Sanger, D. (2011). The transcriptomic responses of the eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular Ecology*, *20*, 1431–1449. https://doi.org/10.1111/j.1365-294X.2011.05018.x

Chen, H. B., & Paul, C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, *12*, 35.

Cheviron, Z. A., Whitehead, A., & Brumfield, R. T. (2008). Transcriptomic variation and plasticity in rufous-collared sparrows (*Zonotrichia*

capensis) along an altitudinal gradient. *Molecular Ecology*, *17*, 4556–4569.

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. k., Tucker, A., Oakley, T. H., … Boore, J. l. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science*, *331*, 555–561. https://doi.org/10.1126/science.1197761

Cook, D. R. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*, 15–18.

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

Danecek, P., Auton, A., Abecasis, G., Banks, E., DePristo, M. A., Handsaker, R. E., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Boer, M. E., Ellers, J., van Gestel, C. A., den Dunnen, J. T., van Straalen, N. M., & Roelofs, D. (2013). Transcriptional responses indicate attenuated oxidative stress in the springtail *Folsomia candida* exposed to mixtures of cadmium and phenanthrene. *Ecotoxicology*, *22*, 619–631. https://doi.org/10.1007/s10646-013-1053-1

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. https://doi.org/10.1016/S1146-609X(02)01145-1

Declerck, S., Cousyn, C., & De Meester, L. (2001). Evidence for local adaptation in neighbouring *Daphnia* populations: A laboratory transplant experiment. *Freshwater Biology*, *46*, 187–198. https://doi.org/10.1046/j.1365-2427.2001.00632.x

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498. https://doi.org/10.1038/ng.806

Desmarais, K. H. (1997). Keeping *Daphnia* out of the surface film with cetyl alcohol. *Journal of Plankton Research*, *19*, 149–154.

Eads, B. D., Andrews, J., & Colbourne, J. K. (2008). Ecological genomics in *Daphnia:* Stress responses and environmental sex determination. *Heredity*, *100*, 184–190. https://doi.org/10.1038/sj.hdy.6800999

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

Ebert, D., Zschokke-Rohringer, C. D., & Carius, H. J. (1998). Within–and between–population variation for resistance of *Daphnia magna* to the bacterial endoparasite *Pasteuria ramosa*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*, 2127–2134.

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. https://doi.org/10.1016/j.cub.2008.01.063

Feulner, P. G., Chain, F. J., Panchal, M., Eizaguirre, C., Kalbe, M., Lenz, T. l., … Bornberg-bauer, E. (2013). Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology*, *22*, 635–649. https://doi.org/10.1111/j.1365-294X.2012.05680.x

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., … Punta, M. (2014). Pfam: The protein families database. *Nucleic Acids Research*, *42*, D222–D230. https://doi.org/10.1093/nar/gkt1223

Fraser, H. B. (2011). Genome-wide approaches to the study of adaptive gene expression evolution. *BioEssays*, *33*, 469–477. https://doi.org/10.1002/bies.201000094

Garbutt, J. S., O'Donoghue, A. J., McTaggart, S. J., Wilson, P. J., & Little, T. J. (2014). The development of pathogen resistance in *Daphnia magna*: Implications for disease spread in age-structured populations. *Journal of Experimental Biology*, *217*, 3929–3934. https://doi.org/10.1242/jeb.111260

Giorello, F. M., Feijoo, M., D'Elía, G., Naya, D. E., Valdez, L., Opazo, J. C., & Lessa, E. P. (2018). An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). *Molecular Ecology*, *27*(16), 3274–3286.

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.

Hamrova, E., Mergeay, J., & Petrusek, A. (2011). Strong differences in the clonal variation of two *Daphnia* species from mountain lakes affected by overwintering strategy. *BMC Evolutionary Biology*, *11*, 231.

Haupt, F., Stockenreiter, M., Baumgartner, M., Boersma, M., & Stibor, H. (2009). *Daphnia* diel vertical migration: Implications beyond zooplankton. *Journal of Plankton Research*, *31*, 515–524. https://doi.org/10.1093/plankt/fbp003

Havel, J. E., Colbourne, J. K., & Hebert, P. D. N. (2000). Reconstructing the history of intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers. *Limnology and Oceanography*, *45*, 1414–1419.

Hayashi, Y., Heckmann, L. H., Callaghan, A., & Sibly, R. M. (2008). Reproduction recovery of the crustacean *Daphnia magna* after chronic exposure to ibuprofen. *Ecotoxicology*, *17*, 246–251. https://doi.org/10.1007/s10646-008-0191-3

Heckmann, L. H., Callaghan, A., Hooper, H. L., et al. (2007). Chronic toxicity of ibuprofen to *Daphnia magna*: Effects on life history traits and population dynamics. *Toxicology Letters*, *172*, 137–145. https://doi.org/10.1016/j.toxlet.2007.06.001

Henning-Lucass, N., Cordellier, M., Streit, B., & Schwenk, K. (2016). Phenotypic plasticity in life-history traits of *Daphnia galeata* in response to temperature - a comparison across clonal lineages separated in time. *Ecology and Evolution*, *6*, 881–891.

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, *173*, 579–588. https://doi.org/10.1086/597611

Herrmann, M. (2017). *Evolutionary genomics in* Daphnia. Mainz, Germany: Universität Koblenz-Landau.

Herrmann, M., Ravindran, S. P., Schwenk, K., & Cordellier, M. (2017a) Data from: Population transcriptomics in *Daphnia*: The role of thermal selection. Dryad Data Repository.

Herrmann, M., Ravindran, S. P., Schwenk, K., & Cordellier, M. (2017b). Population transcriptomics in *Daphnia*: The role of thermal selection. *Molecular Ecology*, *27*, 387–402.

Hodgins, K. A., Yeaman, S., Nurkowski, K. A., Rieseberg, L. H., & Aitken, S. N. (2016). Expression divergence is correlated with sequence evolution but not positive selection in conifers. *Molecular Biology and Evolution*, *33*, 1502–1516. https://doi.org/10.1093/molbev/msw032

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

Jansen, R. C., & Nap, J.-P. (2001). Genetical genomics: The added value from segregation. *Trends in Genetics*, *17*, 388–391. https://doi.org/10.1016/S0168-9525(01)02310-1

Jueterbock, A., Franssen, S. U., Bergmann, N., Gu, J., Coyer, J. A., Reusch, T. B., … Olsen, J. L. (2016). Phylogeographic differentiation versus transcriptomic adaptation to warm temperatures in *Zostera marina*, a globally important seagrass. *Molecular Ecology*, *25*, 5396–5411.

Karhunen, M., Merila, J., Leinonen, T., Cano, J. M., & Ovaskainen, O. (2013). DRIFTSEL: An R package for detecting signals of natural selection in quantitative traits. *Molecular Ecology Resources*, *13*, 746–754.

Karhunen, M., & Ovaskainen, O. (2012). Estimating population-level coancestry coefficients by an admixture F model. *Genetics*, *192*, 609–617. https://doi.org/10.1534/genetics.112.140871

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*, 3059–3066. https://doi.org/10.1093/nar/gkf436

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Keller, L. F. W., & Donald, M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*, 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kesari, R., Lasky, J. R., Villamor, J. G., Des Marais, D. I., Chen, Y.-J., Liu, T.-W., … Verslues, P. E. (2012). Intron-mediated alternative splicing of *Arabidopsis P5CS1* and its association with natural variation in proline and climate adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 9197–9202. https://doi.org/10.1073/pnas.1203433109

Kitano, J., Bolnick, D. I., Beauchamp, D. A., Mazur, M. M., Mori, S., Nakano, T., & Peichel, C. L. (2008). Reverse evolution of armor plates in the threespine stickleback. *Current Biology*, *18*, 769–774. https://doi.org/10.1016/j.cub.2008.04.027

Kudapa, H., Azam, S., Sharpe, A. G., Taran, B., Li, R., Deonovic, B., … Varshney, R. K. (2014). Comprehensive transcriptome assembly of chickpea (*Cicer arietinum L.*) using Sanger and next generation sequencing platforms: Development and applications. *PLoS ONE*, *9*, e86039.

Langfelder & Horvath (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

Langfelder, P., Luo, R., Oldham, M. C., & Horvath, S. (2011). Is my network module preserved and reproducible? *PLoS Computational Biology*, *7*, e1001057.

Larsen, P. F., Nielsen, E. E., Williams, T. D., & Loeschcke, V. (2008). Intraspecific variation in expression of candidate genes for osmoregulation, heme biosynthesis and stress resistance suggests local adaptation in European flounder (*Platichthys flesus*). *Heredity*, *101*, 247–259. https://doi.org/10.1038/hdy.2008.54

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J. Jr, & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*, 2178–2189. https://doi.org/10.1101/gr.1224503

Lopez-Maury, L., Marguerat, S., & Bahler, J. (2008). Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, *9*, 583–593. https://doi.org/10.1038/nrg2398

Louette, G., & De Meester, L. (2004). Rapid colonization of a newly created habitat by cladocerans and the initial build-up of a *Daphnia*-dominated community. *Hydrobiologia*, *513*, 245–249. https://doi.org/10.1023/B:hydr.0000018299.54922.57

Louette, G., Vanoverbeke, J., Ortells, R., & De Meester, L. (2007). The founding mothers: The genetic structure of newly established *Daphnia* populations. *Oikos*, *116*, 728–741.

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

Luikart, G., Allendorf, F. W., Cornuet, J. M., & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, *89*, 238–247. https://doi.org/10.1093/jhered/89.3.238

Macháček, J. (1991). Indirect effect of planktivorous fish on the growth and reproduction of *Daphnia galeata*. *Hydrobiologia*, *225*, 193–197. https://doi.org/10.1007/BF00028397

MacManes, M. D., & Eisen, M. B. (2014). Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus*. *PeerJ*, *2*, e642.

Macnair, M. R. (2002). Within and between population genetic variation for zinc accumulation in *Arabidopsis halleri*. *New Phytologist*, *155*, 59–66. https://doi.org/10.1046/j.1469-8137.2002.00445.x

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

Mergeay, J., Aguilera, X., Declerck, S., Petrusek, A., Huyse, T., & De Meester, L. (2008). The genetic legacy of polyploid Bolivian *Daphnia:* The tropical Andes as a source for the North and South American *D. pulicaria* complex. *Molecular Ecology*, *17*, 1789–1800.

Mills, S., Lunt, D. H., & Gomez, A. (2007). Global isolation by distance despite strong regional phylogeography in a small metazoan. *BMC Evolutionary Biology*, *7*, 225. https://doi.org/10.1186/1471-2148-7-225

Miner, B. E., De Meester, L., Pfrender, M. E., Lampert, W., & Hairston, N. G. Jr (2012). Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings. Biological Sciences*, *279*, 1873–1882.

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., & Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, *41*, e121. https://doi.org/10.1093/nar/gkt263

Munoz, J., Chaturvedi, A., De Meester, L., & Weider, L. J. (2016). Characterization of genome-wide SNPs for the water flea *Daphnia pulicaria* generated by genotyping-by-sequencing (GBS). *Scientific Reports*, *6*, 28569. https://doi.org/10.1038/srep28569

Munoz, J., Gomez, A., Green, A. J., Figuerola, J., Amat, F., & Rico, C. (2008). Phylogeography and local endemism of the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda: Anostraca). *Molecular Ecology*, *17*, 3160–3177.

Okun, N., Mendonca, R., & Mehner, T. (2005). Diel shifts in community composition and feeding of juvenile fishes in the pelagic area of a large shallow lake. *Limnologica – Ecology and Management of Inland Waters*, *35*, 70–77. https://doi.org/10.1016/j.limno.2005.01.005

Oleksiak, M. F., Churchill, G. A., & Crawford, D. L. (2002). Variation in gene expression within and among natural populations. *Nature Genetics*, *32*, 261–266. https://doi.org/10.1038/ng983

Orsini, L., Brown, J. B., Shams Solari, O., Li, D., He, S., Podicheti, R., & De Meester, L. (2017). Early transcriptional response pathways in *Daphnia magna* are coordinated in networks of crustacean-specific genes. *Molecular Ecology*, *27*, 886–897.

Orsini, L., Gilbert, D., Podicheti, R., Jansen, M., Brown, J. B., Solari, O. S., Frilander, M. J. (2016). *Daphnia magna* transcriptome by RNA-Seq across 12 environmental stressors. *Scientific Data*, *3*, 160030. https://doi.org/10.1038/sdata.2016.30

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, *22*, 5983–5999. https://doi.org/10.1111/mec.12561

Ortells, R., Vanoverbeke, J., Louette, G., & De Meester, L. (2013). Colonization of *Daphnia magna* in a newly created pond: Founder effects and secondary immigrants. *Hydrobiologia*, *723*, 167–179. https://doi.org/10.1007/s10750-013-1593-7

Ovaskainen, O., Zheng, C., Karhunen, M., Cano Arias, J. M., & Merilä, J. (2011). A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics*, *189*(2), 621–632.

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

Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420. https://doi.org/10.1093/bioinformatics/btp696

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pietrzak, B., Pijanowska, J., & Dawidowicz, P. (2015). The effect of temperature and kairomone on *Daphnia* escape ability: A simple bioassay. *Hydrobiologia*, *798*, 15–23. https://doi.org/10.1007/s10750-015-2539-z

Poynton, H. C., Zuzow, R., Loguinov, A. V., Perkins, E. J., & Vulpe, C. D. (2008). Gene expression profiling in *Daphnia magna*, part II: Validation of a copper specific gene expression signature with effluent from two copper mines in California. *Environmental Science & Technology*, *42*, 6257–6263.

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rausher, M. D., & Delph, L. F. (2015). Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution*, *69*, 1655–1664.

Reusch, T. B., Veron, A. S., Preuss, C., Weiner, J., Wissler, L., Beck, A., Bornberg-Bauer, E. (2008). Comparative analysis of expressed sequence tag (EST) libraries in the seagrass *Zostera marina* subjected to temperature stress. *Marine Biotechnology*, *10*, 297–309. https://doi.org/10.1007/s10126-007-9065-6

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(6), 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Roberge, C., Guderley, H., & Bernatchez, L. (2007). Genomewide identification of genes under directional selection: Gene transcription Q_{ST} scan in diverging Atlantic salmon subpopulations. *Genetics*, *177*, 1011–1022. https://doi.org/10.1534/genetics.107.073759

Rockman, M. V. (2008). Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature*, *456*, 738–744. https://doi.org/10.1038/nature07633

Roelofs, D., Janssens, T. K., Timmermans, M. J., Nota, B., Mariën, J., Bochdanovits, Z., & VanStraalen, N. M. (2009). Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod *Orchesella cincta*. *Molecular Ecology*, *18*, 3227–3239.

Romero, I. G., Ruvinsky, I., & Gilad, Y. (2012). Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, *13*, 505–516. https://doi.org/10.1038/nrg3229

Routtu, J., & Ebert, D. (2015). Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasites. *Heredity*, *114*, 241–248. https://doi.org/10.1038/hdy.2014.97

Roy Chowdhury, P., Frisch, D., Becker, D., Lopez, J. A., Weider, L. J., Colbourne, J. K., & Jeyasingh, P. D. (2015). Differential transcriptomic responses of ancient and modern *Daphnia* genotypes to phosphorus supply. *Molecular Ecology*, *24*, 123–135.

Roy Chowdhury, P., Lopez, J. A., Weider, L. J., Colbourne, J. K., & Jeyasingh, P. D. (2014). Functional genomics of intraspecific variation in carbon and phosphorus kinetics in *Daphnia*. *Journal of Experimental Zoology. Part A, Ecological Genetics and Physiology*, *321*, 387–398.

Savolainen, O., Lascoux, M., & Merila, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*, 807–820. https://doi.org/10.1038/nrg3522

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., … Friend, S. H. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature*, *422*, 297–302. https://doi.org/10.1038/nature01434

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

Soetaert, A., Vandenbrouck, T., van der Ven, K., Maras, M., van Remortel, P., Blust, R., & De Coen, W. M. (2007). Molecular responses during cadmium-induced stress in *Daphnia magna*: Integration of differential gene expression with higher-level effects. *Aquatic Toxicology*, *83*, 212–222. https://doi.org/10.1016/j.aquatox.2007.04.010

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

St Pierre, S. E., Ponting, L., Stefancsik, R., McQuilton, P., & FlyBase, C. (2014). FlyBase 102–advanced approaches to interrogating FlyBase. *Nucleic Acids Research*, *42*, D780–D788. https://doi.org/10.1093/nar/gkt1092

Stajich, J. E., & Hahn, M. W. (2005). Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution*, *22*, 63–73.

Straile, D., & Walter, G. (1998). Crustacean zooplankton in Lake Constance from 1920 to 1995: Response to eutrophication and re-oligotrophication. *Advances in Limnology*, *53*, 255–274.

Sudheesh, S., Sawbridge, T. I., Cogan, N. O., Kennedy, P., Forster, J. W., & Kaur, S. (2015). *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics*, *16*, 611. https://doi.org/10.1186/s12864-015-1815-7

Taillon Miller, P., Gu, Z., Li, Q., Hillier, L., & Kwok, P. Y. (1998). Overlapping genomic sequence a treasure trove of single nucleotide polymorphisms. *Genome Research*, *8*, 748–754.

Tams, V., Luneburg, J., Seddar, L., Detampel, J. P., & Cordellier, M. (2018). Intraspecific phenotypic variation in life history traits of *Daphnia* galeata populations in response to fish kairomones. *PeerJ*, *6*, e5746.

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

Thielsch, A., Brede, N., Petrusek, A., De Meester, L., & Schwenk, K. (2009). Contribution of cyclic parthenogenesis and colonization history to population structure in *Daphnia*. *Molecular Ecology*, *18*, 1616–1628.

Thielsch, A., Glass, N., Streit, B., De Meester, L., Ortells, R., & Schwenk, K. (2015). Fitness differences and persistent founder effects determine the clonal composition during population build-up in *Daphnia*. *Oikos*, *124*, 620–628.

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.

Vatanparast, M., Shetty, P., Chopra, R., Doyle, J. J., Sathyanarayana, N., & Egan, A. N. (2016). Transcriptome sequencing and marker development in winged bean (*Psophocarpus tetragonolobus*; Leguminosae). *Scientific Reports*, *6*, 29070. https://doi.org/10.1038/srep29070

Ventura, M., Petrusek, A., Miro, A., Hamrová, E., Buñay, D., De Meester, L., & Mergeay, J. (2014). Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential. *Molecular Ecology*, *23*, 1014–1027. https://doi.org/10.1111/mec.12656

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, *150*(3811), 563–565. https://doi.org/10.1038/150563a0

Weisberg, S. F. (2011). *An R companion to applied regression*. Thousand Oaks, CA: Sage.

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, *327*, 343–348. https://doi.org/10.1126/science.1178028

Whitehead, A., & Crawford, D. L. (2006a). Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 5425–5430.

Whitehead, A., & Crawford, D. L. (2006b). Variation within and among species in gene expression: Raw material for evolution. *Molecular Ecology*, *15*, 1197–1211.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

Yampolsky, L. Y., Zeng, E., Lopez, J., Williams, P. J., Dick, K. B., Colbourne, J. K., & Pfrender, M. E. (2014). Functional genomics of acclimation and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics*, *15*. https://doi.org/10.1186/1471-2164-15-859

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

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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ravindran SP, Herrmann M, Cordellier M. Contrasting patterns of divergence at the regulatory and sequence level in European *Daphnia galeata* natural populations. *Ecol Evol*. 2019;9:2487–2504. https://doi.org/10.1002/ece3.4894