bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

3 Suda Parimala Ravindran[1], Maike Herrmann[2] and Mathilde Cordellier[1*]
4 [1]Universität Hamburg, Institute of Zoology, Martin-Luther-King Platz 3, 20146
5 Hamburg, Germany
6 [2]Department of Veterinary Medicine, Paul-Ehrlich-Institut, Paul-Ehrlich-Straße 51-59,
7 63225 Langen, Germany
8
9 *Corresponding author:
10 Martin-Luther-King Platz 3
11 20146 Hamburg, Germany
12 Tel: +49-(0)42838-3933
13 Email: mathilde.cordellier@uni-hamburg.de

14 Running title: Regulatory and sequence divergence in *Daphnia*

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

15 **ABSTRACT**

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

32 **Keywords:** constitutive gene expression, RNA-seq, molecular phenotype, population
33 transcriptomics, DRIFTSEL

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

34 **INTRODUCTION**

35 Natural genetic variation shapes divergence in phenotypic traits and is an
36 important resource for evolutionary processes (Oleksiak *et al.* 2002). Populations
37 respond to environmental variation by genetically adapting to their environments
38 (Hereford 2009; Kawecki & Ebert 2004; Savolainen *et al.* 2013), often showing
39 variations at both gene expression and sequence level across the geographic range
40 of a species. One of the fundamental goals of research in the field of molecular
41 evolution is to resolve the evolutionary processes driving the rise and maintenance of
42 expression and sequence polymorphisms behind this variation. Revealing their effect
43 on an organism's fitness thereby aids to understand the genetic basis of local
44 adaptation (MacManes & Eisen 2014). Gene expression patterns link genotypes and
45 phenotypes, sometimes called a "molecular phenotype", and as such is an important
46 component in local adaptation processes (Lopez-Maury *et al.* 2008). --Several
47 studies have reported the testing of different populations exposed to different
48 treatments and examining their transcriptional response, for example in springtails
49 (*Folsomia* (De Boer *et al.* 2013) and *Orchesella* (Roelofs *et al.* 2009)), oyster
50 (*Crassostrea virginica*; Chapman *et al.* 2011; Chapman *et al.* 2009), sparrows
51 (*Zonotrichia capensis*; Cheviron *et al.* 2008), flounder (*Platichthys flesus*; Larsen *et*
52 *al.* 2008), and seagrass (*Zostera marina*; Jueterbock *et al.* 2016; Reusch *et al.* 2008),
53 thereby identifying candidate genes involved in local adaptation. Gene expression
54 variation can be highly heritable (Brem & Kruglyak 2005; Schadt *et al.* 2003;
55 Whitehead & Crawford 2006b). Moreover, constitutive gene expression patterns also
56 differ within- and among- natural populations (e.g., Roberge *et al.* 2007; Whitehead &
57 Crawford 2006a), strongly suggesting that standing variation in constitutive gene

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 expression is shaped by local adaptation. Natural selection acts immediately on
59 newly arisen variation (in contrast to adaptation observed from standing genetic
60 variation) as there are neutral and slightly deleterious variations preserved in a
61 population, which may become beneficial upon changes in selection regimes (Barrett
62 & Schluter 2008). After a sudden change of environment, standing variation can
63 contribute to fast adaptation (Feulner *et al.* 2013; Kitano *et al.* 2008). Identifying
64 allelic/genetic variants underlying differences in expression profiles can be helpful in
65 hypothesizing gene functions (Jansen & Nap 2001; Kesari *et al.* 2012; Rockman
66 2008). Although prior knowledge of the specific loci is not a prerequisite to learn
67 about adaptive processes in most cases, identification of genetic features underlying
68 local adaptation is critical in answering fundamental questions about natural selection
69 (Rausher & Delph 2015).

70 Geneticvariation within and among populations is strongly influenced by their
71 colonization history, and the demographic changes following the primary
72 establishment of a population. Population sizes may vary after colonization across
73 the species based on environmental factors and further colonization (Böndel *et al.*
74 2015). Colonization events depend on dispersal ability, and dispersal rates strongly
75 differ from gene flow estimates in several species (De Meester *et al.* 2002). This is
76 particularly evident in freshwater zooplankton species, where several studies suggest
77 a high potential for dispersal when populations rapidly colonize new habitats and
78 spread invasively (Havel *et al.* 2000; Louette & De Meester 2004; Mergeay *et al.*
79 2008). However, genetic studies show that the observed rate of gene flow is much
80 lower than would be expected in organisms with high dispersal potential (Boileau *et*
81 *al.* 1992; De Meester *et al.* 2002; Thielsch *et al.* 2009). This ambiguity between

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

82 dispersal potential and rate of gene flow can be explained by founder effects (Boileau
83 *et al.* 1992) complemented by local adaptation; resulting in monopolization of
84 resources by local populations (De Meester *et al.* 2002). This process leads to the
85 impression that population genetic variation correlates with the colonization patterns
86 (Orsini *et al.* 2013).

87 Amongst freshwater zooplankton species, the water flea *Daphnia* is the best
88 studied and has been an important model for ecology, population genetics,
89 evolutionary biology, and toxicology (Ebert 2005). This genus belongs to the order
90 Cladocera and has attracted scientific interest since the 17th century (Desmarais
91 1997). It inhabits most types of freshwater habitats and includes more than 100
92 known species of freshwater plankton organisms (Ebert 2005). *Daphnia* make an
93 interesting subject of investigation in comparative functional genomics (Eads *et al.*
94 2008). Apart from the fact that *Daphnia* species have an appropriate size for being
95 used in laboratory cultures, they are easy to cultivate and have short generation
96 times. Because of their clonal mode of reproduction, *Daphnia* are highly suited for
97 quantitative genetic studies, which can enhance our understanding of their
98 evolutionary ecology.

99 Genetic variation has been reported for numerous traits in *Daphnia,* such as
100 life history traits (e.g., Henning-Lucass *et al.* 2016), vertical migration (e.g., Haupt *et*
101 *al.* 2009), fish escape behavior (e.g., Pietrzak *et al.* 2015), resistance against
102 parasites (e.g., Routtu & Ebert 2015) and immune responses (e.g., Garbutt *et al.*
103 2014). Furthermore, it was shown that responses to many chemical stressors such
104 as phosphorus (Roy Chowdhury *et al.* 2015; Roy Chowdhury *et al.* 2014), copper
105 (Poynton *et al.* 2008), cadmium (Soetaert *et al.* 2007) and pharmaceutical products

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

106 like ibuprofen (Hayashi *et al.* 2008; Heckmann *et al.* 2007) have a genetic basis as
107 well. Within- and between-population comparisons in *Daphnia* have been conducted
108 extensively using varied environmental perturbations and providing evidences for
109 local adaptation (for e.g., Barata *et al.* 2002; Declerck *et al.* 2001; Ebert *et al.* 1998;
110 Spitze 1993). Although various aspects like phylogeography, functional morphology,
111 physiology and life history evolution have been in the limelight of *Daphnia* research
112 for several decades (Eads *et al.* 2008), *Daphnia* genomics investigations have begun
113 only in the last decade with the availability of the *Daphnia pulex* genome (Colbourne
114 *et al.* 2011). A considerable number of studies (for e.g.: Bento *et al.* 2017; Miner *et*
115 *al.* 2012; Orsini *et al.* 2016; Yampolsky *et al.* 2014) on biotic and abiotic factors have
116 been carried out showing how *Daphnia* respond to environmental perturbations by
117 changes in gene expression. However, little is known about the intra-specific
118 variability at the gene expression level in *Daphnia*, since the above-mentioned
119 studies focused on stressor driven responses using a reduced number of genotypes.
120 To sum up, elucidating the mechanisms by which natural selection acts on
121 gene expression evolutionremains a challenge (e.g.: Fraser 2011; Romero *et al.*
122 2012). Unraveling the relative consequences of drift versus natural selection on gene
123 expression profiles plays an important role in understanding species divergence and
124 local adaptation. The studies listed above provided evidence for gene expression
125 variation correlated with many environmental factors in *Daphnia*. However,
126 knowledge about the variation in constitutive gene expression structure within and
127 among population is lacking.
128 In the present study on *Daphnia galeata,* sampled from four different lakes in
129 Europe, we conducted a large scale RNA-seq study in the absence of any laboratory

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

130 induced environmental stressor. Using transcriptome data, we quantified the
131 constitutive expression profiles and performed a sequence analysis of the four
132 populations. We addressed the following questions: (i) Are there differences in gene
133 expression profiles between the four populations? (ii) How is the observed variation
134 explained by the different levels of organization, i.e., genotype and population? (iii)
135 Do the observed differences in expression profiles result from genetic drift or
136 selection? (iv) What is the role of genetic drift and/or natural selection in shaping
137 sequence variation? (v) What are the functional roles of the transcripts?

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

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

149 **METHODS**

150 **Sampling and RNA collection**

151 A set of *D. galeata* resting stages (ephippia) was collected from the sediment
152 of four lakes: Jordán Reservoir (hereafter, Pop.J) in Czech Republic, Müggelsee in
153 Germany (hereafter, Pop.M), Lake Constance (hereafter, Pop.LC) at the border
154 between Germany, Switzerland and Austria, and Greifensee (hereafter, Pop.G) in
155 Switzerland. These ephippia were hatched under laboratory conditions (see Henning-
156 Lucass *et al.* 2016 for hatching conditions) and the hatchlings were used to establish
157 clonal lines in a laboratory setting. The species identity was checked by sequencing
158 a fragment of the *12S* mitochondrial locus and 10 microsatellite markers (Multiplex 2
159 comprising the loci *Dgm109, Dp196, Dp281, Dp512, SwiD1, SwiD10, SwiD12,*
160 *SwiD14, SwiD15, SwiD2*), following protocols by Taylor *et al.* (1996) and Yin *et al.*
161 (2010) respectively.

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

169 **RNA preparation and sequencing**

170 Total RNA was extracted following a modified phenol/chloroform protocol and
171 further processed using the RNeasy kit (Qiagen, Hilden, Germany). The total RNA
172 was eluted in RNAse free water and the concentration and quality (RNA integrity

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

173 number and phenol) were checked using a NanoDrop spectrophotometer (Thermo
174 Scientific, Wilmington, DE, USA) and a Bioanalyzer 2100 (Agilent Technologies,
175 Santa Clara, CA, USA). The 72 total RNA samples were sent to the company GATC
176 (Konstanz, Germany) for library preparation and sequencing. Following reverse
177 transcription and cDNA construction using random primers, 50bp single-end (SE)
178 reads were sequenced on an Illumina HiSeq 2000 (San Diego, CA, USA), with. To
179 avoid block effects and confounding effects in the downstream analysis, we used a
180 completely randomized design; each library was sequenced on at least two different
181 lanes, on a total of nine lanes. Detailed information can be found in Table S1.

182 **Quality trimming, mapping and read counts**

183 All reads with ambiguous bases (Ns) were removed before trimming. Bases
184 with a phred score below 20 were trimmed at the 3' and 5' ends. Reads shorter than
185 45 bp after trimming were discarded. All trimming steps were conducted using locally
186 installed version of Galaxy at the Gene Center in Munich, Germany.

187 Trimmed reads were mapped to the reference *D. galeata* transcriptome
188 (Huylmans *et al.* 2016; available from NCBI: https://www.ncbi.nlm.nih.gov, GenBank
189 ID: HAFN00000000.1) using NextGenMap (Sedlazeck *et al.* 2013) with increased
190 sensitivity (-i 0.8 –kmer-skip 0 -s 0.0). Read counts were obtained from the SAM files
191 using a custom python script (available upon request) and discarding ambiguously
192 mapped reads. The raw count table was analyzed in R (R Development Core Team
193 2008) using the package DESeq2 (Love *et al.* 2014). Normalization was done with
194 size factor procedure. Standard differential analysis steps of DESeq2 such as
195 estimation of dispersion and negative binomial GLM fitting were applied. The count
196 outliers were automatically detected using Cook's distance, which is a measure of

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

197 how much the fitted coefficients would change if an individual sample was removed
198 (Cook 1977). Principal Component Analysis (PCA) was performed to visualize the
199 clustering of biological replicates and clonal lines.

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

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

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

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

218

219

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

220 We searched for transcripts for which the identified differential expression
221 could not be explained by phylogenetic distance and genetic drift alone. To identify
222 signals of possible selection, we used the approach of Ovaskainen *et al.* (2011)
223 implemented in the R package DRIFTSEL 2.1.2 (Karhunen *et al.* 2013), considering
224 expression of every single transcript as a trait. To perform this analysis, we made use
225 of the microsatellite data and normalized read count values. Allele frequencies were
226 obtained from microsatellite data collected in a previous study, independently from
227 the species identification step outlined above. Microsatellite data of 30-40 resting
228 eggs also sampled from the same sediment layers the resurrected clonal lines come
229 from was obtained from a study by Herrmann (2017). Briefly, eleven microsatellite
230 loci were analyzed for each individual according to the protocol published by Thielsch
231 *et al.* (2009). Primers for all loci were multiplexed and PCR was performed using the
232 Type-it Microsatellite PCR Kit (Qiagen, Hilden, Germany). Alleles were recorded
233 manually and allelic frequencies were calculated with GenAlEx (Peakall & Smouse
234 2012).

235 Using microsatellite allelic frequencies, the coancestry coefficients by
236 admixture F model was calculated using “do.all” function implemented in the RAFM
237 package (Karhunen & Ovaskainen 2012). We ran a total of 200,000 iterations with
238 thinning at an interval of 1,000 and discarded the first 1,000 iterations as ‘burn-in’.
239 The output was a list which contained samples from the posterior distributions of
240 allele frequencies. Values from the posterior coancestry matrix, ’theta’, were used as
241 input for the Metropolis-Hastings (MH) algorithm along with the normalized read
242 counts for DETs as implemented in DRIFTSEL. We ran a total of 5,000 iterations with
243 thinning at 1,000 samples and discarded the first 100 iterations as burn-in. The

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

244 output of MH algorithm was a matrix of posterior of subpopulation effects (pop.ef),
245 used to estimate the H.test values. The H.test describes whether the population
246 means correlate with the genetic data more than it would be expected on basis of
247 shared evolutionary history. Large H-values imply that the populations are more
248 locally adapted than expected by chance.

249

250 **Intra and inter-population variation**

251 To quantify the respective contributions of the factors "genotype" and
252 "population" to the observed variation in gene expression profiles, we performed an
253 analysis of variance (ANOVA) test in R on the normalized read counts obtained from
254 DESeq2. To correct for multiple testing, p_{adj}-values were calculated for each
255 transcript using the Benjamini-Hochberg procedure.

256 **Variant calling and filtering**

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

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

276 **Neutrality statistics**

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

284 Results from LOSITAN (Antao *et al.* 2008) outlier tests were obtained from
285 Herrmann *et al.* (2017b) to identify loci under selection (see Table S4).

286

287 **Heterozygosity and mutation frequencies**

288 The heterozygosity values for the final SNP dataset were calculated with
289 VCFtools (Danecek *et al.* 2011). The ratio between the expected heterozygosity (H_E)
290 and observed heterozygosity (H_O) was calculated based on available SNP
291 information and plots were created using ggplot2 (Wickham 2009) in R.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

292 **Sequence vs. regulatory variation**

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

300

301 **Annotation and functional analysis**

302 To functionally annotate the *D. galeata* transcripts, a local sequence alignment
303 using blastn (Altschul *et al.* 1990) against the nr database (downloaded Feb. 2015
304 via ftp://ftp.ncbi.nlm.nih.gov/blast/db/) was performed. Hits with an eval ≤ 0 and
305 identity ≥ 50% were considered. Additionally, protein domain annotations and
306 orthoMCL (Li *et al.* 2003) results were obtained from (Huylmans *et al.* 2016) . Briefly,
307 a search was made for all three *Daphnia* species (*D. pulex, D. magna* and *D.*
308 *galeata*) using PfamScan (version 1.5) to look into the Pfam A database (version
309 27.0; Finn *et al.* 2014) together with hmmer3 (version 3.1b; Mistry *et al.* 2013). In
310 order to identify orthologs and be able to compare it to other arthropod species,
311 orthoMCL was used to cluster the amino acid sequences of *D galeata*, *D. pulex*
312 (version JGI060905; Colbourne *et al.* 2011), *D. magna* (version 7; *Daphnia*
313 Genomics Consortium 2015), as well as *Drosophila melanogaster* (version 5.56; St
314 Pierre *et al.* 2014) and *Nasonia vitripennis* (version 1.2; Werren *et al.* 2010) into
315 orthologous groups and determine the inparalogs. Pie charts representing the

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

316 number of hits obtained for all transcripts and DETs were created using the plotrix
317 package (Lemon 2006) in R.
318 We classified the orthoMCL clusters into the following categories:
319 (a) Clusters that contain only *D. galeata*-specific transcripts
320 (b) Clusters that are shared between *D. galeata* and *D. pulex*
321 (c) Clusters that are shared between *D. galeata* and *D. magna*
322 (d) Clusters that are shared between *D. galeata*, *D. pulex* and *D. magna*
323 (*Daphnia*-specific)
324 (e) Clusters that are shared between *D. galeata* and other arthropods
325 (*D. melanogaster* and *N. vitripennis*)
326 (f) Clusters that are shared among all five analyzed species (*Daphnia* and both
327 insects)

328 **Inparalogs and misassemblies**

329 To assess whether *D. galeata* DETs in an orthologous group are “inparalogs”,
330 isoforms or the result of misassembly, we computed the pairwise sequence
331 divergence for those orthoMCL clusters containing DETs from at least two different
332 populations. Since each significantly differentially expressed transcript was assigned
333 as a DET only to the population in which it was upregulated the most, clusters
334 containing more than one DET most likely contained DETs from different populations.
335 Based on the number of populations within their orthoMCL cluster, the DETs were
336 classified into the categories: “1Pop”, “2Pop”, “3Pop” and “4Pop”, and unclustered
337 DETs were categorized as “0Pop”. 0Pop and 1Pop DETs were excluded from further
338 analysis. In total, there were 716 orthoMCL clusters that contained DETs from at
339 least two populations. Pairwise alignments of the amino acid sequences in each

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

340 orthologous group were performed using the iterative refinement method
341 incorporating local pairwise alignment information (L-INS-i) in MAFFT (Katoh *et al.*
342 2002). We then used EMBOSS tranalign (Rice *et al.* 2000) to generate alignments of
343 nucleic acid coding regions translated from aligned protein sequences. Pairwise
344 genetic divergence was computed with 'dist.dna' function implemented in the ape
345 package (Paradis *et al.* 2004) in R, using the Kimura-2-parameters model with
346 gamma correction. We used an arbitrary cut-off value of 2 to distinguish inparalogs
347 from misassembled sequences.

348

349 **Gene Ontology (GO) enrichment analysis**

350 DETs with a H.value ≥ 0.95 (DRIFTSEL result) and transcripts with a nonzero
351 D value in each of the four populations (Tajima's D result) were analyzed with
352 „topGO“ (Alexa & Rahnenfuhrer 2016) in R, using a custom GO annotation for *D.*
353 *galeata.* GO terms enriched in the transcripts of interest in each population from each
354 analysis (DRIFTSEL and Tajima's D) were identified using the 'weight01' algorithm
355 for all three ontologies, namely: molecular function, cellular component and biological
356 processes. We used a Fisher test and those GO terms with a classicFisher value ≤
357 0.05 were considered to be enriched for each ontology in each population. A multiple
358 testing procedure was not applied as the *p*-values returned by the 'weight01'
359 algorithm are interpreted to be corrected and might exclude "true" annotations (Alexa
360 & Rahnenfuhrer 2016).

361

362 **Weighted Gene Co-expression network analysis**

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

363 To gain insights into the population-specific regulatory patterns of transcripts in
364 *D. galeata*, we performed a weighted gene co-expression network analysis with
365 WGCNA (Langfelder & Horvath 2008) using the variance stabilized normalized read
366 counts obtained in DESeq2 analysis. Transcripts and samples that had lower
367 expression values were excluded from every population using the
368 'goodSampleGenes' function in WGCNA and used for downstream analysis. In total,
369 32375 transcripts were used for the construction of gene co-expression networks. To
370 identify population-specific co-expression modules (i.e., clusters of highly correlated
371 transcripts), a network was first built using the full dataset (i.e., with samples and
372 transcripts from all populations) and one network for each population using
373 expression values specific to all genotype and biological replicates. The population
374 specific network was compared to the reference network and an adjacency matrix
375 was calculated. Clusters were identified using the WGCNA Topological Overlap
376 Matrices (TOM). For every transcript and module detected automatically, WGCNA
377 assigns a color based on the module membership (MM) value. An MM value is a
378 measure of module membership which is obtained by correlating its gene expression
379 profiles with module eigengene (i.e., the first principal component of a given module).
380 For example, if a transcript has an MMred value close to ±1, the transcript is
381 assigned to the red module (Langfelder & Horvath 2008). Each module is assigned a
382 color based on the module size: 'turquoise' denotes the largest module, blue next,
383 followed by brown, green, yellow and so on. The color 'grey' is reserved for
384 unassigned transcripts (Langfelder & Horvath 2008). Similarly, the module 'gold'
385 consists of 1000 randomly selected transcripts that represent a sample of the whole

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

386 network and statistical measures have no meaning for this module (Langfelder &
387 Horvath 2008).

388 After obtaining the module definitions from each comparison, we assessed
389 how well our modules in the reference network are preserved in the population
390 specific networks using the 'modulePreservation' function, which outputs a single Z-
391 score summary. The higher the Z-score, the more preserved a module is between
392 the reference and population-specific network. A module was deemed to be
393 preserved if the Z-score value was above 10, an arbitrary value deemed suitable by
394 Langfelder *et al.* (2011).
395

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

396 **RESULTS**

397 **Sequencing results and mapping statistics**

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

408 **Differential expression**

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

418 After conducting pairwise contrast analyses with DESeq2, we identified
419 transcripts exclusively upregulated for each population when compared to all others

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

420 (p_{adj} ≤ 0.05; thereafter DETs). In total, 10,820 of 32,903 transcripts (~33%) showed
421 significant expression differences in pairwise comparisons between populations. Of
422 all ~33000 transcripts, 9.6%, 8.1%, 7.2% and 7.8% were population specific DETs for
423 the populations Pop.G, Pop.J, Pop.LC and Pop.M, respectively (Figure 1b).
424

425 **Role of natural selection on transcript expression levels**

426 The DRIFTSEL multivariate approach was used to identify transcripts for
427 which the observed differential expression could not be explained by phylogenetic
428 distance and genetic drift alone; the alternative explanation being that the observed
429 divergence would be attributable to selection and therefore possibly to local
430 adaptation events. In total, 48% of 10820 differentially expressed transcripts showed
431 greater differential expression than expected under neutrality (H-value ≥ 0.95, Figure
432 1b and Table S2), indicating that the observed pattern is due to local adaptation for
433 these transcripts. Pop.LC had the highest number of DETs deviating from the neutral
434 expectations (67% transcripts out of 2381), followed by Pop.G (~44% out of 3163),
435 Pop.J (~49% out of 2679), and Pop.M (35% out of 2597).

436 **Expression variation among individuals and populations**

437 The statistical significance of difference between group means of expression
438 values was assessed with two ANOVAs for each transcript, one grouping the data
439 according to genotypes and the other according to populations. For 1,897
440 transcripts, the means were statistically significantly different between populations
441 but not between genotypes. The reverse was true for 7,546 transcripts. For 15,323
442 transcripts, the factors 'genotype' and 'population' explained the observed variation in

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

443 gene expression. The remaining 8,137 transcripts had no significant p_{adj}-values in
444 either of the ANOVAs.

445 **Sequence based divergence**

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

450 A PCA was carried out based on a matrix of all biallelic SNP sites to illustrate
451 the population structure among the four populations. Although PC1 explained the
452 maximum variance (12%) (Figure 2a) and four distinct clusters corresponding to the
453 populations were seen against PC2. PC2 and PC3 each explained 8% of the
454 variance (Figure S2). PC2 clearly separated the genotypes belonging to Pop.J from
455 the remainder of the data (Figure 2a).

456 **Heterozygosity**

457 The observed heterozygosity values ranged from -0.19 to 0.22 for genotypes
458 from Pop.G, from 0.08 to 0.17 for those from Pop.J, from 0.07 to 0.15 within Pop.LC,
459 and from -0.06 to 0.32 for those belonging to Pop.M (Figure 2b). Within Pop.G, four
460 out of the six genotypes exhibited an observed heterozygosity lower than the
461 expected heterozygosity. In Pop.J, Pop.LC, and Pop.M (except genotype M9
462 therein), the observed heterozygosity exceeded the expected heterozygosity;
463 implying higher genetic variability in these genotypes.

464 **Sequence evolution**

465 To assess the respective contributions of random and non-random evolutionary
466 events on DNA sequence divergence, we calculated the Tajima's D statistic for each

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

467 transcript in the four populations. After phasing, we obtained 13,006 transcripts
468 containing SNPs. Pop.LC had the highest number of transcripts (32.21%) with a
469 negative D value (D < 0; $p \leq 0.05$) followed by Pop.G (30.45% transcripts), Pop.M
470 (29.58% transcripts) and Pop.J (29.31% transcripts). Much fewer transcripts were
471 found to have a significant positive Tajima's D value (Table 2): 1.26% transcripts in
472 Pop.M, 1.20% transcripts in Pop.G, 1.13% transcripts in Pop.J and 0.66% transcripts
473 in Pop.LC (Table S3).

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

478 **Sequence vs. regulatory variation**

479 The proportion of transcripts identified to be candidates for local adaptation at
480 both sequence and regulatory level were visualized using a flow diagram (Figure 3).
481 Among the 10,820 transcripts identified to be differentially expressed, ~46% showed
482 signs of selection at the regulatory level according to DRIFTSEL. Of these, ~15%
483 were identified as outliers under balancing and/or diversifying selection in LOSITAN.
484 About 26% of these outliers had a significantly negative or positive Tajima's D value
485 in at least one population, which might be attributed to selection but can also stem
486 from other evolutionary processes such as population growth, reduction or
487 subdivision, bottleneck events and migration.

488

489 **Functional annotation**

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

490 Of all transcripts, 66.5% had a BLAST hit to the nr database with an identity
491 ≥ 50% and eval ≤ 0; 91.4% transcripts of these BLAST hits shared homology with
492 other *Daphnia* species. Among the DETs, 70.4% met this criterion (Supplementary
493 Figure S3a, Table S5), and 92.3% of them were homologous to *Daphnia* sequences.
494 We were able to predict domains for ~50% of our transcripts. Among the
495 DETs, a slightly higher proportion of transcripts, ~53%, had known protein domains
496 (Supplementary Figure S3b, Table S5).
497 For identifying *Daphnia*-specific orthologs and those that share orthology with
498 other arthropods, the orthoMCL data was classified into six categories (as described
499 in the Methods section). 3,058 orthology clusters (of which 1,735 clusters contained
500 DETs) were containing exclusively *D. galeata* transcripts, 985 clusters (of which 543
501 clusters contained DETs) contained only *D. galeata* transcripts and *D. pulex* genes,
502 651 clusters (including 224 DETs) contained only *D. galeata* and *D. magna*
503 transcripts. 3336 orthoMCL clusters (of which 1239 clusters contained *D.galeata*
504 DETs) contained all three *Daphnia* species used in the analysis. Furthermore, 12
505 clusters (4 clusters containing DETs) were containing *D. galeata* transcripts along
506 with two other arthropods (*D. melanogaster* and *N. vitripennis*). In total, 4657 clusters
507 (1586 clusters containing DETs) contained transcripts/genes for all five species
508 (three *Daphnia* species and two insects) used in the present study (Supplementary
509 Figure S3c, Table S5).
510

511 **Assessment of assembly artefacts and inparalogs**

512 In total, 3,325 DETs belonged to the 0Pop category (Figure 4a), 5,574 DETs
513 were exclusively occurring in orthoMCL clusters without DETs from different

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

514 populations (1Pop). This vast majority was thus not further analyzed with regard to
515 paralogy and assembly artefacts. The remaining 1,921 DETs were co-occurring with
516 DETs from other populations in 716 orthoMCL clusters. Sequence divergence was
517 calculated for every DET pair that co-occurred in a cluster. The divergence values
518 ranged from 0.0 to 12.0 (Figure 4b). We cannot exclude that divergence values
519 greater than 2 between sequence pairs arose from misassemblies. However, 16,752
520 sequence pairs (belonging to 671 clusters) had a divergence lower than our arbitrary
521 threshold of 2, indicating that the transcripts were highly similar in their sequence and
522 thus might constitute inparalogs or alternative transcripts for a gene. In this case,
523 only genomic data would allow placing the transcripts and eventually assessing their
524 status.

525 **Gene Ontology enrichment analysis**

526 GO enrichment analysis was performed on the candidate transcripts as
527 identified from DRIFTSEL (Hvalue ≥ 0.95) and Tajima's D analyses. We observed an
528 enrichment for several metabolic processes such as ATP binding, DNA binding,
529 microtubule binding, transporter activities and signaling pathways (Table S6) in both
530 analyses in all population-specific sets. Specifically, in Pop.G, DRIFTSEL and
531 Tajima's D analysis had five GO terms in common, in Pop.J, they had one GO term
532 in common, in Pop.LC they had four GO terms in common and in Pop.M, they had
533 seven GO terms in common.

534

535 **Weighted Gene Co-expression network analysis**

536 The WGCNA on 32,375 transcripts identified 29 co-expression modules
537 (Figure 5) in the reference network (see Methods). We observed varying numbers of

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

538 modules and transcripts clustered in each population-specific network (Table S7a-d).
539 However, after assessing the conserved modules, where each population-specific
540 network was compared to the reference network, 24 modules (out of 29) were well
541 conserved (Zscore ≥ 10) among the populations. The conserved modules included
542 10,256 transcripts altogether, which is about 31% of all transcripts in *D. galeata,* with
543 the largest module, 'turquoise' including 2,857 transcripts. Two modules (grey and
544 gold) with uncharacterized and random transcripts contained 16,600 and 1000
545 transcripts, respectively. These results are consistent with the gene expression
546 analysis which showed little differences between the populations.

547 **DISCUSSION**

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

559 ***Population divergence at the sequence level***

560 SNPs became the absolute marker of choice for molecular genetic analysis as the
561 mining of polymorphisms is the cheapest source for genetic variability (Taillon Miller

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

562 *et al.* 1998). Our PCA analysis on SNP data revealed four clear population clusters
563 and our results are in agreement with a highly structured population model across the
564 transcriptome. Although two of the genotypes (G1.6 and G1.7) from Pop.G were
565 located outside the Pop.G cluster in the PCA plot, the populations were clearly
566 distinguished and corresponded to the four lakes sampled. This pattern might be the
567 result of several non-exclusive phenomena: initial founder effects, isolation-by-
568 distance and genetic drift, and natural divergent selection, since the studied
569 populations originate from lakes located in different ecoregions.

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

582 All the phenomena cited above have an impact on population structure across
583 the genome, and might mask highly diverging loci resulting from natural selection.
584 We assessed patterns of divergence at the sequence level through neutrality tests
585 (Tajima's *D*). This suggested that all populations of *Daphnia* examined in this study

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

586 had a substantial amount (~48% transcripts) of loci with an excess of low frequency
587 polymorphisms (i.e., D < 0) relative to the neutral expectation. This pattern may result
588 from positive selection, a bottleneck, or population expansion. It is consistent with
589 previous observations in *Daphnia* from Lake Greifensee and Lake Constance (Brede
590 *et al.* 2009) and crustacean zooplankton from Lake Constance (Straile 1998) which
591 have all undergone historical bottleneck events. Similarly, Lake Müggelsee, a large
592 shallow lake, has undergone severe bottlenecks due to increased turbidity and
593 because vegetation disappeared almost completely after the 1960s (Okun *et al.*
594 2005). One other explanation for the excess of rare alleles is selection against
595 genotypes carrying deleterious alleles.

596 Although a high frequency of rare polymorphisms was observed in our
597 analysis, there were few transcripts (~1.7% transcripts) that had a lower frequency of
598 rare alleles (D > 0) in the four populations; indicating that some loci are either under
599 balancing selection (where heterozygous genotypes are favored) or under
600 diversifying selection (where genotypes carrying the less common alleles are
601 favored). A lower frequency of rare alleles also occurs if there is a recent population
602 admixture (Stajich & Hahn 2005). This argument is consistent with our heterozygosity
603 measures. Under the Hardy-Weinberg equilibrium, genotypes G2.1 and G3.1 from
604 Pop.G, all genotypes in Pop.J and Pop.LC, and all genotypes except M9 in Pop.M
605 show that the observed heterozygosity is greater than the expected heterozygosity,
606 which is an indication of higher genetic variability and population admixture. Most of
607 the genotypes in population G, as well as M9, have a much lower heterozygosity.
608 Such low heterozygosity patterns at the individual level can be attributed to
609 inbreeding (Keller 2002), but also due to a lack of variation in the source population,

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

610 either caused by a small founder population size or a severe bottleneck during
611 population history (Luikart *et al.* 1998). While genotype M9 from Müggelsee might be
612 an exception, the pattern observed in Greifensee could be the consequence of
613 inbreeding and/or low genetic variability in this population, either resulting from
614 previous bottlenecks, or a reduced number of "founding mothers". Further, the
615 ecology and growth dynamics of *Daphnia* populations might exacerbate the founder
616 effects. After an initial hatching phase from the resting eggs bank and exponential
617 population growth in the spring, clonal selection occurs throughout the growing
618 season (Vanoverbeke & De Meester 1997). Therefore, it is possible that only a few
619 clonal lines contribute to the resting eggs population each year. However, while a
620 reduced number of clonal lines might contribute to the yearly "archiving" of genetic
621 diversity; two processes counteract the immediate diversity loss. First, the spring
622 recruitment doesn't only rely on eggs from the previous year but rather on a mixture
623 (Vanoverbeke & De Meester 1997), and might even integrate overwintering clones in
624 larger permanent lakes (but see Yin *et al.* 2014 for an overview). Second, clonal
625 erosion doesn't affect the same genotypes every year, leading to year-to-year
626 heterogeneity, such as the one observed in the long term study by Griebel *et al.*
627 (2016). Clonal erosion thus doesn't necessarily lead to a downward spiral of genetic
628 diversity loss, and the high stochasticity of both clonal selection and hatching ensure
629 a preservation of the genetic diversity in every habitat.

630 ***Gene expression variability and signals of selection***

631 While the patterns observed at the sequence level tends to support the role of
632 genetic drift, founder and monopolization effects in shaping the observed patterns,
633 the results of our gene expression analysis delivered a mixed message. This was

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

634 evident in the PCA based on the gene expression data, where no distinct clusters
635 corresponding to populations are clearly visible. This observation was consistent with
636 our network co-expression analysis which showed that the identified modules are
637 conserved in all populations (Figure 5), with a few exceptions. The analysis of
638 variance confirms this finding, with a relatively low number of transcripts for which the
639 mean read counts differs significantly between populations and not between
640 genotypes. While studies on differential expression in *Gliricidia sepium* (Chalmers et
641 al. 1992) and *Arabidopisis halleri* (Macnair 2002) have observed substantial between
642 population variances at the gene expression level, our results are consistent with
643 several studies, for example, on fish (*Fundulus heteroclitus*; Whitehead & Crawford
644 2006a) and snails (*Melanoides tuberculata*; Facon *et al.* 2008) which showed large
645 within-population variation. Additionally, numerous studies on life-history traits in
646 *Daphnia* also report very high intrapopulation variability (Beckerman *et al.* 2010;
647 Castro *et al.* 2007; Cousyn *et al.* 2001; Macháček 1991). A common garden
648 experiment conducted on the very same clonal lines also showed a higher
649 phenotypic variability within populations than among populations (V. Tams, personal
650 communication). Finally, the observed relative homogeneity in the gene expression
651 profiles might be the consequence of high selective pressure on transcription
652 regulation or canalization (Waddington 1942). Such canalization allows for storage of
653 cryptic genetic variation that would be uncovered in stress response assessments.
654 However, our experimental setup was designed to avoid stress, and transcriptome
655 characterization of the same genotypes under conditions mimicking predation,
656 parasite or food stress, for example, might reveal a greater divergence between the
657 populations.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

658 Comparisons of the gene expression profiles for the four populations revealed
659 a fair number ~8% of *D. galeata* transcripts to be significantly exclusively upregulated
660 in one given population compared to all others. Although all populations showed
661 similar numbers of differentially upregulated transcripts, when considering those
662 which are probably under directional selection, the picture changed. After applying
663 the DRIFTSEL approach, Pop.LC had the highest number of transcripts directionally
664 selected based on their expression levels and Pop.M had the lowest number. Pop.G
665 and Pop.J had nearly similar numbers of transcripts under directional selection. This
666 discrepancy in the number of transcripts that are differentially expressed and those
667 presumably under directional selection can partially be explained by parallel
668 adaptation to contrasting environments. A study on adaptive differentiation in
669 seagrass (Jueterbock *et al.* 2016) that compared Northern and Southern seagrass
670 samples under thermal stress showed that natural selection was the most
671 straightforward explanation for nearly 1% of all differentially expressed genes. For
672 other genes that were differentially expressed in the seagrass study, parallel
673 adaptation to different habitats was observed along both the American and European
674 thermal clines.

675 -***Sequence vs. regulatory variation in* Daphnia galeata**

676 Correlating expression profiles with sequence divergence helps to identify
677 transcripts that are potentially under the influence of local adaptation at both gene
678 expression and sequence level. Linking gene expression profiles with sequence
679 polymorphisms and their associated functions aids in understanding the genetic
680 basis of adaptation as seen in the desert adapted mouse (*Peromyscus eremicus*;
681 MacManes & Eisen 2014) and in the Patagonian olive mouse (*Abrothrix olivacea*;

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

682 Giorello *et al.* 2018). Our results revealed ~30% of the transcripts to share
683 divergence at both sequence and regulatory level (Figure 3). There are two possible
684 explanations for the observed differences in sequence and regulatory level variation
685 (Hodgins *et al.* 2016). The first is that there is an increase in the rate of fixation due to
686 transcripts under positive selection and divergence in expression patterns. For
687 example, variation in gene expression might lead to selection for sequence variation
688 to improve the functional role of the transcript in its altered role (Hodgins *et al.* 2016).
689 A second explanation is that the differentially expressed transcripts may experience
690 reduced negative selection in one or all four populations. For instance, higher
691 transcript expression is associated with greater negative selection. Hence a reduction
692 in transcript expression in one population compared to others may be accompanied
693 by relaxation of selection in that population.

694 GO enrichment analysis on the candidates identified at the sequence (Tajima's D)
695 and expression (DRIFTSEL) level were enriched for metabolic and cellular
696 processes. These findings suggest that there may be a hierarchical activation of
697 general mechanisms of stress responses at the metabolic and cellular level. This
698 observation is concordant to another study (Orsini *et al.* 2017) on *D. magna.* In this
699 study, *D. magna* were subjected to several environmental perturbations and the GO
700 enrichment analysis revealed a general stress response rather than ontologies
701 specific to local adaptation. Since the present study is without any laboratory induced
702 stressor, further studies in *Daphnia* subjected to one or multiple environmental
703 stressors would be helpful in pinpointing stress specific responses. Further, no GO
704 term annotation was available for ~31% of the transcripts, and we can therefore no
705 reach conclusive results. This highlights the need for new and complementary

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

706 resources for *Daphnia* genomics research, and a general improvement of the existing
707 annotation.
708

709 ***Gene annotation and evaluation of inparalogs***

710 Gene annotation is quite challenging in organisms lacking reference genomes, and
711 functional annotation then relies on the availability of transcriptomic sequences from
712 the closest available taxon. In this study we were able to annotate 66.5% of the
713 transcripts using BLAST analysis (Supplementary Figure S3a). However, many of the
714 transcripts were homologous to a *D. pulex* "hypothetical protein", likely because (i)
715 they are similar in function to non-coding regions or pseudogenes or (ii) novel coding
716 transcripts that are yet to be functionally characterized (Vatanparast *et al.* 2016).
717 Furthermore, we were able to predict domains for 80% of the transcripts using Pfam
718 analysis (Supplementary Figure S3b, Supplementary Table S4). Our orthoMCL
719 results (Supplementary Figure S3c, Supplementary Table S4) showed that several
720 (~45%) of the *D. galeata* transcripts were orthologous to one or all species of
721 *Daphnia* used for comparison, indicating that the genes/transcripts have all evolved
722 from a common *Daphnia*-specific ancestral gene via speciation. In addition to this,
723 ~25% of *Daphnia* genes/transcripts are orthologous to two insect species (*D.*
724 *melanogaster* and *N. vitripennis*). Our level of unannotated transcripts is similar to
725 results reported from other organisms lacking extensive genomic resources, for
726 example, from plants like field pea (*Pisum sativum;* Sudheesh *et al.* 2015), chick pea
727 (*Cicer arietinum;* Kudapa *et al.* 2014), and winged bean (*Psophocarpus*
728 *tetragonolobus*; Vatanparast *et al.* 2016). This limited our interpretation of the
729 functional role of *Daphnia* transcripts and thereby their associations to known

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

730 ecological stressors. A second issue raised when lacking a reference genome is that
731 it might be difficult to tease apart inparalogs created by duplication events, isoforms
732 and even misassemblies; leading to an artificially inflated number of similar
733 sequences for each distinct gene in the transcript set. Only ~18% of the population
734 specific DETs had one or more putative paralogs also identified as differentially
735 expressed in at least one other population. For DETs from two or more populations
736 that co-occurred in orthoMCL clusters, we were able to distinguish between actual
737 paralogs (transcript pairs that had a sequence divergence value > 2, Figure 4b) and
738 transcripts with sequence divergence value < 2.Genomic information is now required
739 for this species in order to accurately assign transcripts to genes and correctly
740 assess whether two different populations might indeed express different gene copies
741 with similar functions
742
743 **FUTURE DIRECTIONS AND CONCLUSIONS**
744 In summary, we described here an approach that combines both transcriptomic
745 expression profiles and sequence information to understand local adaptation in
746 *D. galeata*. Although the set of transcripts contributing to population divergence at the
747 sequence and the expression level differ, both levels constitute alternative routes for
748 responding to selection pressures (Pai *et al.* 2015); showing that these transcripts
749 can contribute to local adaptation and paving way for future research. From our
750 functional analysis, it was evident that most of our transcripts were *Daphnia* specific
751 although they had hypothetical functions. To understand the function of the
752 hypothetical transcripts in *D. galeata* and their response to environmental
753 perturbations, a comparative approach using the gene expression data from

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

754 numerous other *Daphnia* studies should be used. Although we noticed correlations
755 between expression patterns and sequence divergence for the *D. galeata* transcripts,
756 we lack genomic and phylogenetic information. This information may help "bridge the
757 gap" for understanding the relative roles of positive or negative selection in driving
758 coding sequence and gene expression divergence.
759
760 *Data accessibility*
761 The raw sequence reads used for this study as well as the experimental set up for
762 the analysis of differentially expressed genes are available on ArrayExpress
763 (https://www.ebi.ac.uk/arrayexpress; Accession no.: E-MTAB-6144).
764 The raw read counts used as input for differential transcript expression , results for
765 the pairwise contrast analysis conducted in DESEq2 ,and the number of variants per
766 sample before and after filtering, number of variant sites per transcript are all
767 available on DRYAD in Tables S10, S11 and S3-S4, respectively
768 (https://datadryad.org//resource/doi:10.5061/dryad.p85m5). The VCF file will be
769 made available on European Variant Archive (EVA) and accession numbers will be
770 updated.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

771 *Supporting Information*

772 Supporting information File 1:

773

774 Table S1: Library preparation and sequencing information along with principal
775 component coordinates for the first three axes as obtained from gene expression
776 analysis.
777 Table S2: DRIFTSEL values for the differentially expressed transcripts.
778 Table S3: Population-wise Tajima's D values.
779 Table S4: LOSITAN outlier test values to identify loci under selection.
780 Table S5: Functional annotation for candidate transcripts of local adaptation.
781 Table S6a-c: Population specific GO enrichment terms using DRIFTSEL and
782 Tajima's D analysis.
783 Table S7a-d: Number of transcripts clustered in each module as detected by
784 WGCNA for PopG, PopJ, PopLC and PopM.
785 Supporting information File 2:
786 Figure S1 Gene expression PCA for the first three principal components
787 Figure S2 SNP PCA for the first three principal components
788 Figure S3a-c Pie charts showing functional annotation using BLAST, Pfam
789 and orthoMCL analysis.
790

791 *Author's contributions*

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

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

796 *Acknowledgements*

797 We would like to thank A. Jueterbock and O. Ovaskainen for suggestions and help
798 on the DRIFTSEL implementation, and three anonymous reviewers for their
799 comments on an earlier version of the manuscript. This study was financially
800 supported by an individual grant from the Volkswagen Stiftung (to MC). This work
801 benefits from and contributes to the *Daphnia* Genomics Consortium.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

802 Figure legends:
803 Figure 1: Gene expression patterns. (a) Gene expression PCA of the four sampled
804 populations: Pop.G (Lake Greifensee), Pop.J (Jordan Reservoir), Pop.LC (Lake
805 Constance) and Pop.M (Müggelsee). Percentages on the X- and Y- axis indicate the
806 percentage of variance explained by each principal component. (b) Venn diagram
807 illustrating the number of differentially expressed transcripts (DET) between the four
808 populations. Numbers in brackets indicate the number of transcripts deviating from
809 the neutral expectations according to the DRIFTSEL analysis.

810 Figure 2: SNP patterns and heterozygosity. (A) SNP PCA of the four sampled
811 populations: Pop.G (Lake Greifensee); Pop.J (Jordan reservoir), Pop.LC (Lake
812 Constance) and Pop.M (Müggelsee). Percentages on the X- and Y-axis indicate the
813 percentage of variance explained by each principal component. (B) Barplot
814 illustrating the heterozygosity values for each genotype.

815 Figure 3: Flow diagram representing the proportion of transcripts that are candidates
816 for local adaptation at the regulatory and sequence level. Each analysis or “step” is
817 represented by a vertical group of black rectangle bars, called nodes. The colored
818 areas linking the nodes are called “flows”. The **DESeq2** step contains four nodes:
819 PopG (yellow), PopJ (black), PopLC (pink) and PopM (green), which represent the
820 number of transcripts specifically upregulated in each of the four populations as
821 identified by DESeq2 analysis. The **DRIFTSEL** step contains 2 nodes: ‘H.value ≤
822 0.95’ (grey) and ‘H.value ≥ 0.95’ (purple). The **LOSITAN** step contains 5 nodes: ‘NC’
823 (grey) with transcripts without LOSITAN result (not calculated); ‘noOL’ (grey):
824 transcripts where none of the SNPs in a transcript were identified as outliers; ‘Bal’
825 (cyan), transcripts containing at least one SNP that is under balancing selection; ‘Div’
826 (pink) transcripts containing at least one SNP under diversifying selection; and
827 ‘BalDiv’ (pale green), transcripts containing SNPs that are under both balancing and
828 diversifying selection. The **Tajima’s D** step contains 8 nodes. Each node classifies
829 the transcripts according to the obtained Tajima’s D values. ‘AllNeg’ means that
830 transcripts have a negative D value in all four populations; ‘AllPos’ means that
831 transcripts have a positive D value in all four populations; ‘AllNonSig’ means
832 transcripts have non-significant D values in all four populations; ‘NegNonsig’ means
833 transcripts in the four populations have either a negative D value or a nonsignificant
834 D value; ‘PosNonsig’ means transcripts in the four populations have either a positive
835 D value or a nonsignificant D value; ‘PosNeg’ means transcripts in the four
836 populations have either a positive or negative D value; ‘PosNegNonsig’ means
837 transcripts in the four populations have either a positive or negative or an insignificant
838 D value.

839

840 Figure 4: Differentiating misassembly from inparalogs. (a) Barplot showing the
841 number of DETs co-occurring with DETs from other populations within an orthoMCL
842 cluster. 0Pop refers to DETs not assigned to an orthoMCL cluster. 1Pop, 2Pop, 3Pop
843 and 4Pop refer to DETs found in orthoMCL clusters containing at least one, two,
844 three and four population(s) respectively. (b) Histogram of pairwise sequence
845 divergence values calculated for all *D. galeata* sequences co-occurring in an
846 orthoMCL cluster belonging to 2Pop, 3Pop and 4Pop categories.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

847 :

848 Figure 5: Cluster dendrogram of transcripts for the reference network in *Daphnia*
849 *galeata*, with dissimilarity based on the topological overlap matrix (TOM). The co-
850 expression modules are colored in an arbitrary way by the WGCNA package, and the
851 size of the bar is proportional to the number of transcripts in the module. The right
852 hand side grid represents the module conservation in each population. Modules with
853 a Z-score ≤ 10 are shown in white and modules with a Z-score ≥10 are colored in
854 dark grey.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

855 References
856

857 Akey JM (2003) The effect of single nucleotide polymorphism identification strategies
858 on estimates of linkage disequilibrium. *Molecular Biology and Evolution* **20**,
859 232-242.
860 Alexa A, Rahnenfuhrer J (2016) topGO: Enrichment analysis for gene ontology.
861 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment
862 search tool. *Journal of Molecular Biology* **215**, 403-410.
863 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a
864 workbench to detect molecular adaptation based on a F_{ST}-outlier method.
865 *BMC Bioinformatics* **9**, 323.
866 Barata C, Baird DJ, Mitchell SE, Soares AMVM (2002) Among- and within-population
867 variability in tolerance to cadmium stress in natural populations of *Daphnia*
868 *magna*. *Environmental Toxicology and Chemistry* **21**, 1058-1064.
869 Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. *Trends in*
870 *Ecology & Evolution* **23**, 38-44.
871 Beckerman AP, Rodgers GM, Dennis SR (2010) The reaction norm of size and age
872 at maturity under multiple predator risk. *Journal of Animal Ecology* **79**, 1069-
873 1076.
874 Benjamini YH, Yosef (1995) Controlling the false discovery rate: a practical and
875 powerful approach to multiple testing. *Journal of the Royal Statistical Society.*
876 *Series B* **57**, 289-300.
877 Bento G, Routtu J, Fields PD*, et al.* (2017) The genetic basis of resistance and
878 matching-allele interactions of a host-parasite system: The *Daphnia magna-*
879 *Pasteuria ramosa* model. *PLoS Genetics* **13**, e1006596.
880 Boileau MG, Hebert PDN, Schwartz SS (1992) Non-equilibrium gene frequency
881 divergence: persistent founder effects in natural populations. *Journal of*
882 *Evolutionary Biology* **5**, 25-39.
883 Böndel KB, Lainer H, Nosenko T*, et al.* (2015) North-South colonization associated
884 with local adaptation of the wild tomato species *Solanum chilense*. *Molecular*
885 *Biology and Evolution* **32**, 2932-2943.
886 Brede N, Sandrock C, Straile D*, et al.* (2009) The impact of human-made ecological
887 changes on the genetic architecture of *Daphnia* species. *Proceedings of the*
888 *National Academy of Sciences of the United States of America* **106**, 4758-
889 4763.
890 Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene
891 expression traits in yeast. *Proceedings of the National Academy of Sciences*
892 *of the United States of America* **102**, 1572-1577.
893 Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and
894 missing-data inference for whole-genome association studies by use of
895 localized haplotype clustering. *American Journal of Human Genetics* **81**, 1084-
896 1097.
897 Castro BB, Consciência S, Gonçalves F (2007) Life history responses of *Daphnia*
898 *longispina* to mosquitofish (*Gambusia holbrooki)* and pumpkinseed (*Lepomis*
899 *gibbosus*) kairomones. *Hydrobiologia* **594**, 165-174.
900 Chalmers KJ, Waugh R, Sprent JI, Simons AJ, Powell W (1992) Detection of genetic
901 variation between and within populations of *Gliricidia sepium* and *G. maculata*
902 using RAPD markers. *Heredity* **69**.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

903 Chapman RW, Mancia A, Beal M*, et al.* (2011) The transcriptomic responses of the
904 eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular*
905 *Ecology* **20**, 1431-1449.
906 Chapman RW, Mancia A, Beal M*, et al.* (2009) A transcriptomic analysis of land-use
907 impacts on the oyster, *Crassostrea virginica,* in the South Atlantic bight.
908 *Molecular Ecology* **18**, 2415-2425.
909 Chen HB, Paul C (2011) VennDiagram: a package for the generation of highly-
910 customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**.
911 Cheviron ZA, Whitehead A, Brumfield RT (2008) Transcriptomic variation and
912 plasticity in rufous-collared sparrows (*Zonotrichia capensis)* along an
913 altitudinal gradient. *Molecular Ecology* **17**, 4556-4569.
914 Colbourne JK, Pfrender ME, Gilbert D*, et al.* (2011) The ecoresponsive genome of
915 *Daphnia pulex*. *Science* **331**, 555-561.
916 Cook DR (1977) Detection of influential observation in linear regression.
917 *Technometrics* **19**, 15-18.
918 Cousyn C, De Meester L, Colbourne JK*, et al.* (2001) Rapid, local adaptation of
919 zooplankton behavior to changes in predation pressure in the absence of
920 neutral genetic changes. *Proceedings of the National Academy of Sciences of*
921 *the United States of America* **98**, 6256-6260.
922 Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools.
923 *Bioinformatics* **27**, 2156-2158.
924 De Boer ME, Ellers J, van Gestel CA*, et al.* (2013) Transcriptional responses indicate
925 attenuated oxidative stress in the springtail *Folsomia candida* exposed to
926 mixtures of cadmium and phenanthrene. *Ecotoxicology* **22**, 619-631.
927 De Meester L, Gomez A, Okamura B, Schwenk K (2002) The Monopolization
928 Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta*
929 *Oecologica* **23**, 121-135.
930 Declerck S, Cousyn C, De Meester L (2001) Evidence for local adaptation in
931 neighbouring *Daphnia* populations: a laboratory transplant experiment.
932 *Freshwater Biology* **46**, 187-198.
933 DePristo MA, Banks E, Poplin R*, et al.* (2011) A framework for variation discovery
934 and genotyping using next-generation DNA sequencing data. *Nature Genetics*
935 **43**, 491-498.
936 Desmarais KH (1997) Keeping *Daphnia* out of the surface film with cetyl alcohol.
937 *Journal of Plankton Research* **19**, 149-154.
938 Eads BD, Andrews J, Colbourne JK (2008) Ecological genomics in *Daphnia:* stress
939 responses and environmental sex determination. *Heredity (Edinb)* **100**, 184-
940 190.
941 Ebert D (2005) *Ecology, epidemiology and evolution of parasitism in Daphnia*.
942 Ebert D, Zschokke-Rohringer CD, Carius HJ (1998) Within–and between–population
943 variation for resistance of *Daphnia magna* to the bacterial endoparasite
944 *Pasteuria ramosa*. *Proceedings of the Royal Society of London. Series B:*
945 *Biological Sciences* **265**, 2127-2134.
946 Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-
947 history strategies within invasive populations by way of multiple introductions.
948 *Current Biology* **18**, 363-367.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

949 Feulner PG, Chain FJ, Panchal M*, et al.* (2013) Genome-wide patterns of standing
950 genetic variation in a marine population of three-spined sticklebacks.
951 *Molecular Ecology* **22**, 635-649.
952 Finn RD, Bateman A, Clements J*, et al.* (2014) Pfam: the protein families database.
953 *Nucleic Acids Research* **42**, D222-230.
954 Fraser HB (2011) Genome-wide approaches to the study of adaptive gene
955 expression evolution. *Bioessays* **33**, 469-477.
956 Garbutt JS, O'Donoghue AJ, McTaggart SJ, Wilson PJ, Little TJ (2014) The
957 development of pathogen resistance in *Daphnia magna*: implications for
958 disease spread in age-structured populations. *Journal of Experimental Biology*
959 **217**, 3929-3934.
960 Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in*
961 *Genetics* **21**, 616-623.
962 Giorello FM, Feijoo M, D'Elía G*, et al.* (2018) An association between differential
963 expression and genetic divergence in the Patagonian olive mouse (*Abrothrix*
964 *olivacea*). *Molecular Ecology*.
965 Griebel J, Giessler S, Yin M, Wolinska J (2016) Parental and hybrid *Daphnia* from the
966 *D. longispina* complex: long-term dynamics in genetic structure and
967 significance of overwintering modes. *Journal of Evolutionary Biology* **29**, 810-
968 823.
969 Hamrova E, Mergeay J, Petrusek A (2011) Strong differences in the clonal variation
970 of two *Daphnia* species from mountain lakes affected by overwintering
971 strategy. *BMC Evolutionary Biology* **11**.
972 Haupt F, Stockenreiter M, Baumgartner M, Boersma M, Stibor H (2009) *Daphnia* diel
973 vertical migration: implications beyond zooplankton. *Journal of Plankton*
974 *Research* **31**, 515-524.
975 Havel JE, Colbourne JK, Hebert PDN (2000) Reconstructing the history of
976 intercontinental dispersal in *Daphnia lumholtzi* by use of genetic markers.
977 *Limnology and Oceanography* **45**, 1414-1419.
978 Hayashi Y, Heckmann LH, Callaghan A, Sibly RM (2008) Reproduction recovery of
979 the crustacean *Daphnia magna* after chronic exposure to ibuprofen.
980 *Ecotoxicology* **17**, 246-251.
981 Hayes B, Laerdahl JK, Lien S*, et al.* (2007) An extensive resource of single
982 nucleotide polymorphism markers associated with Atlantic salmon (*Salmo*
983 *salar*) expressed sequences. *Aquaculture* **265**, 82-90.
984 Heckmann LH, Callaghan A, Hooper HL*, et al.* (2007) Chronic toxicity of ibuprofen to
985 *Daphnia magna*: Effects on life history traits and population dynamics.
986 *Toxicology Letters* **172**, 137-145.
987 Henning-Lucass N, Cordellier M, Streit B, Schwenk K (2016) Phenotypic plasticity in
988 life-history traits of *Daphnia galeata* in response to temperature - a
989 comparison across clonal lineages separated in time. *Ecology and Evolution*
990 **6**, 881-891.
991 Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The*
992 *American Naturalist* **173**, 579-588.
993 Herrmann M (2017) *Evolutionary Genomics in Daphnia*, Universität Koblenz-Landau.
994 Herrmann M, Ravindran SP, Schwenk K, Cordellier M (2017a) Data from: Population
995 transcriptomics in *Daphnia*: the role of thermal selection. Dryad Data
996 Repository.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

997 Herrmann M, Ravindran SP, Schwenk K, Cordellier M (2017b) Population
998 transcriptomics in *Daphnia*: the role of thermal selection. *Molecular Ecology*.
999 Hodgins KA, Yeaman S, Nurkowski KA, Rieseberg LH, Aitken SN (2016) Expression
1000 divergence is correlated with sequence evolution but not positive selection in
1001 conifers. *Molecular Biology and Evolution* **33**, 1502-1516.
1002 Huylmans AK, Lopez Ezquerra A, Parsch J, Cordellier M (2016) De novo
1003 transcriptome assembly and sex-biased gene expression in the cyclical
1004 parthenogenetic *Daphnia galeata*. *Genome Biology and Evolution* **8**, 3120-
1005 3139.
1006 Jalving R, Vant Slot R, van Oost B (2004) Chicken single nucleotide polymorphism
1007 identification and selection for genetic mapping. *Poultry Science* **83**, 1925-
1008 1931.
1009 Jansen RC, Nap J-P (2001) Genetical genomics: the added value from segregation.
1010 *Trends in Genetics* **17**, 388-391.
1011 Jueterbock A, Franssen SU, Bergmann N, *et al.* (2016) Phylogeographic
1012 differentiation versus transcriptomic adaptation to warm temperatures in
1013 *Zostera marina*, a globally important seagrass. *Molecular Ecology* **25**, 5396-
1014 5411.
1015 Karhunen M, Merila J, Leinonen T, Cano JM, Ovaskainen O (2013) DRIFTSEL: an R
1016 package for detecting signals of natural selection in quantitative traits.
1017 *Molecular Ecology Resources* **13**, 746-754.
1018 Karhunen M, Ovaskainen O (2012) Estimating population-level coancestry
1019 coefficients by an admixture F model. *Genetics* **192**, 609-617.
1020 Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid
1021 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids*
1022 *Research* **30**, 3059-3066.
1023 Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**,
1024 1225-1241.
1025 Keller LFW, Donald M. (2002) Inbreeding effects in wild populations. *Trends in*
1026 *Ecology & Evolution* **17**, 230-241.
1027 Kesari R, Lasky JR, Villamor JG*, et al.* (2012) Intron-mediated alternative splicing of
1028 *Arabidopsis P5CS1* and its association with natural variation in proline and
1029 climate adaptation. *Proceedings of the National Academy of Sciences of the*
1030 *United States of America* **109**, 9197-9202.
1031 Kitano J, Bolnick DI, Beauchamp DA*, et al.* (2008) Reverse evolution of armor plates
1032 in the threespine stickleback. *Current Biology* **18**, 769-774.
1033 Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary fisherian life-history
1034 evolution in small salmonid populations. *Nature* **419**, 824-826.
1035 Kudapa H, Azam S, Sharpe AG*, et al.* (2014) Comprehensive transcriptome
1036 assembly of Chickpea (*Cicer arietinum L.*) using Sanger and next generation
1037 sequencing platforms: development and applications. *PLoS One* **9**, e86039.
1038 Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved
1039 and reproducible? *PLoS Comput Biol* **7**, e1001057.
1040 Larsen PF, Nielsen EE, Williams TD, Loeschcke V (2008) Intraspecific variation in
1041 expression of candidate genes for osmoregulation, heme biosynthesis and
1042 stress resistance suggests local adaptation in European flounder (*Platichthys*
1043 *flesus*). *Heredity (Edinb)* **101**, 247-259.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1044 Leinonen T, McCairns RJ, O'Hara RB, Merila J (2013) Q_{ST}-F_{ST} comparisons:
1045 evolutionary and ecological insights from genomic heterogeneity. *Nature*
1046 *Reviews Genetics* **14**, 179-190.
1047 Lemon J (2006) Plotrix: a package in the red light district of R. **6**, 8-12.
1048 Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format
1049 and SAMtools. *Bioinformatics* **25**, 2078-2079.
1050 Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups
1051 for eukaryotic genomes. *Genome Res* **13**, 2178-2189.
1052 Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing
1053 environments: from rapid responses to evolutionary adaptation. *Nature*
1054 *Reviews Genetics* **9**, 583-593.
1055 Louette G, De Meester L (2004) Rapid colonization of a newly created habitat by
1056 cladocerans and the initial build-up of a *Daphnia*-dominated community.
1057 *Hydrobiologia* **513**, 245-249.
1058 Louette G, Vanoverbeke J, Ortells R, De Meester L (2007) The founding mothers: the
1059 genetic structure of newly established *Daphnia* populations. *Oikos* **116**, 728-
1060 741.
1061 Love MI, Huber W, Anders S (2014) Differential analysis of count data – the DESeq2
1062 package. *bioRxiv (2014).*
1063 Luikart GA, FW, Cornuet J-M, Sherwin W (1998) Distortion of allele frequency
1064 distributions provides a test for recent population bottlenecks. *Journal of*
1065 *Heredity* **89**, 238-247.
1066 Macháček J (1991) Indirect effect of planktivorous fish on the growth and
1067 reproduction of *Daphnia galeata*. *Hydrobiologia* **225**, 193-197.
1068 MacManes MD, Eisen MB (2014) Characterization of the transcriptome, nucleotide
1069 sequence polymorphism, and natural selection in the desert adapted mouse
1070 *Peromyscus eremicus*. *PeerJ* **2**, e642.
1071 Macnair MR (2002) Within and between population genetic variation for zinc
1072 accumulation in *Arabidopsis halleri*. *New Phytologist* **155**, 59-66.
1073 McKenna A, Hanna M, Banks E*, et al.* (2010) The Genome Analysis Toolkit: a
1074 MapReduce framework for analyzing next-generation DNA sequencing data.
1075 *Genome Research* **20**, 1297-1303.
1076 Mergeay J, Aguilera X, Declerck S*, et al.* (2008) The genetic legacy of polyploid
1077 Bolivian *Daphnia:* the tropical Andes as a source for the North and South
1078 American *D. pulicaria* complex. *Molecular Ecology* **17**, 1789-1800.
1079 Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and
1080 quantitative traits. *Journal of Evolutionary Biology* **14**, 892-903.
1081 Michalak P, Minkov I, Helin A*, et al.* (2001) Genetic evidence for adaptation-driven
1082 incipient speciation of *Drosophila melanogaste*r along a microclimatic contrast
1083 in "Evolution Canyon," Israel. *Proceedings of the National Academy of*
1084 *Sciences of the United States of America* **98**, 13195-13200.
1085 Mills S, Lunt DH, Gomez A (2007) Global isolation by distance despite strong
1086 regional phylogeography in a small metazoan. *BMC Evolutionary Biology* **7**,
1087 225.
1088 Miner BE, De Meester L, Pfrender ME, Lampert W, Hairston NG, Jr. (2012) Linking
1089 genes to communities and ecosystems: *Daphnia* as an ecogenomic model.
1090 *Proceedings. Biological sciences* **279**, 1873-1882.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1091 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology
1092 search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic*
1093 *Acids Research* **41**, e121.
1094 Munoz J, Chaturvedi A, De Meester L, Weider LJ (2016) Characterization of
1095 genome-wide SNPs for the water flea *Daphnia pulicaria* generated by
1096 genotyping-by-sequencing (GBS). *Scientific Reports* **6**, 28569.
1097 Munoz J, Gomez A, Green AJ*, et al.* (2008) Phylogeography and local endemism of
1098 the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda:
1099 Anostraca). *Molecular Ecology* **17**, 3160-3177.
1100 Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004) Common pattern of
1101 evolution of gene expression level and protein sequence in *Drosophila*.
1102 *Molecular Biology and Evolution* **21**, 1308-1317.
1103 Okun N, Mendonca R, Mehner T (2005) Diel shifts in community composition and
1104 feeding of juvenile fishes in the pelagic area of a large shallow lake.
1105 *Limnologica - Ecology and Management of Inland Waters* **35**, 70-77.
1106 Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within
1107 and among natural populations. *Nature Genetics* **32**, 261-266.
1108 Orsini L, Brown JB, Shams Solari O*, et al.* (2017) Early transcriptional response
1109 pathways in *Daphnia magna* are coordinated in networks of crustacean-
1110 specific genes. *Mol Ecol*.
1111 Orsini L, Gilbert D, Podicheti R*, et al.* (2016) *Daphnia magna* transcriptome by RNA-
1112 Seq across 12 environmental stressors. *Scientific Data* **3**, 160030.
1113 Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of
1114 population genetic differentiation in the wild: isolation by dispersal limitation,
1115 isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**,
1116 5983-5999.
1117 Ortells R, Vanoverbeke J, Louette G, De Meester L (2013) Colonization of *Daphnia*
1118 *magna* in a newly created pond: founder effects and secondary immigrants.
1119 *Hydrobiologia* **723**, 167-179.
1120 Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends in Genetics*
1121 **29**, 66-73.
1122 Pai AA, Pritchard JK, Gilad Y (2015) The genetic and mechanistic basis for variation
1123 in gene regulation. *PLoS Genetics* **11**.
1124 Paradis E (2010) pegas: an R package for population genetics with an integrated-
1125 modular approach. *Bioinformatics* **26**, 419-420.
1126 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and
1127 evolution in R language. *Bioinformatics* **20**, 289-290.
1128 Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population
1129 genetic software for teaching and research--an update. *Bioinformatics* **28**,
1130 2537-2539.
1131 Pietrzak B, Pijanowska J, Dawidowicz P (2015) The effect of temperature and
1132 kairomone on *Daphnia* escape ability: a simple bioassay. *Hydrobiologia* **798**,
1133 15-23.
1134 Poynton HC, Zuzow R, Loguinov AV, Perkins EJ, Vulpe CD (2008) Gene expression
1135 profiling in *Daphnia magna,* part II: validation of a copper specific gene
1136 expression signature with effluent from two copper mines in California.
1137 *Environmental Science & Technology* **42**, 6257-6263.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1138 R Development Core Team (2008) R: A language and environment for statistical
1139 computing.
1140 Rausher MD, Delph LF (2015) Commentary: When does understanding phenotypic
1141 evolution require identification of the underlying genes? *Evolution* **69**, 1655-
1142 1664.
1143 Reusch TB, Veron AS, Preuss C*, et al.* (2008) Comparative analysis of expressed
1144 sequence tag (EST) libraries in the seagrass *Zostera marina* subjected to
1145 temperature stress. *Marine Biotechnology* **10**, 297-309.
1146 Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology
1147 Open Software Suite. **16**.
1148 Roberge C, Guderley H, Bernatchez L (2007) Genomewide identification of genes
1149 under directional selection: gene transcription Q_{ST} scan in diverging Atlantic
1150 salmon subpopulations. *Genetics* **177**, 1011-1022.
1151 Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural
1152 genetic variation. *Nature* **456**, 738-744.
1153 Roelofs D, Janssens TK, Timmermans MJ*, et al.* (2009) Adaptive differences in gene
1154 expression associated with heavy metal tolerance in the soil arthropod
1155 *Orchesella cincta*. *Molecular Ecology* **18**, 3227-3239.
1156 Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and
1157 the evolution of gene regulation. *Nature Reviews Genetics* **13**, 505-516.
1158 Routtu J, Ebert D (2015) Genetic architecture of resistance in *Daphnia* hosts against
1159 two species of host-specific parasites. *Heredity (Edinb)* **114**, 241-248.
1160 Roy Chowdhury P, Frisch D, Becker D*, et al.* (2015) Differential transcriptomic
1161 responses of ancient and modern *Daphnia* genotypes to phosphorus supply.
1162 *Molecular Ecology* **24**, 123-135.
1163 Roy Chowdhury P, Lopez JA, Weider LJ, Colbourne JK, Jeyasingh PD (2014)
1164 Functional genomics of intraspecific variation in carbon and phosphorus
1165 kinetics in *Daphnia*. *Journal of Experimental Zoology. Part A, Ecological*
1166 *Genetics and Physiology* **321**, 387-398.
1167 Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation.
1168 *Nature Reviews Genetics* **14**, 807-820.
1169 Schadt EE, Monks SA, Drake TA*, et al.* (2003) Genetics of gene expression
1170 surveyed in maize, mouse and man. *Nature* **422**, 297-302.
1171 Sedlazeck FJ, Rescheneder P, von Haeseler A (2013) NextGenMap: fast and
1172 accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**,
1173 2790-2791.
1174 Skelly DA, Ronald J, Akey JM (2009) Inherited variation in gene expression. *Annual*
1175 *Review of Genomics and Human Genetics* **10**, 313-332.
1176 Soetaert A, Vandenbrouck T, van der Ven K*, et al.* (2007) Molecular responses
1177 during cadmium-induced stress in *Daphnia magna*: integration of differential
1178 gene expression with higher-level effects. *Aquatic Toxicology* **83**, 212-222.
1179 Spitze K (1993) Population structure in *Daphnia obtusa* quantitative genetic and
1180 allozymic variation. *Genetics*, 367-374.
1181 St Pierre SE, Ponting L, Stefancsik R, McQuilton P, FlyBase C (2014) FlyBase 102--
1182 advanced approaches to interrogating FlyBase. *Nucleic Acids Research* **42**,
1183 D780-788.
1184 Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection
1185 in human history. *Mol Biol Evol* **22**, 63-73.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1186 Straile DG, Walter (1998) Crustacean zooplankton in Lake Constance from 1920 to
1187 1995: Response to eutrophication and re-oligotrophication. *Advances in*
1188 *Limnology* **53**, 255-274.
1189 Sudheesh S, Sawbridge TI, Cogan NO*, et al.* (2015) *De novo* assembly and
1190 characterisation of the field pea transcriptome using RNA-Seq. *BMC*
1191 *Genomics* **16**, 611.
1192 Taillon Miller P, Gu Z, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequence
1193 a treasure trove of single nucleotide polymorphisms. *Genome Research* **8**,
1194 748-754.
1195 Taylor DJ, Hebert PDN, Colbourne JK (1996) Phylogenetics and evolution of the
1196 *Daphnia longispina* group (Crustacea) based on 12S rDNA sequence and
1197 allozyme variation. *Molecular Phylogenetics and Evolution* **5**, 495-510.
1198 Thielsch A, Brede N, Petrusek A, de Meester L, Schwenk K (2009) Contribution of
1199 cyclic parthenogenesis and colonization history to population structure in
1200 *Daphnia*. *Molecular Ecology* **18**, 1616-1628.
1201 Thielsch A, Glass N, Streit B*, et al.* (2015) Fitness differences and persistent founder
1202 effects determine the clonal composition during population build-up in
1203 *Daphnia*. *Oikos* **124**, 620-628.
1204 Vanoverbeke J, De Meester L (1997) Among-populational genetic differentiation in
1205 the cyclical parthenogen *Daphnia magna* (Crustacea, Anomopoda) and its
1206 relation to geographic distance and clonal diversity. *Hydrobiologia* **360**, 135-
1207 142.
1208 Vatanparast M, Shetty P, Chopra R*, et al.* (2016) Transcriptome sequencing and
1209 marker development in winged bean (*Psophocarpus tetragonolobus*;
1210 Leguminosae). *Scientific Reports* **6**, 29070.
1211 Ventura M, Petrusek A, Miro A*, et al.* (2014) Local and regional founder effects in
1212 lake zooplankton persist after thousands of years despite high dispersal
1213 potential. *Molecular Ecology* **23**, 1014-1027.
1214 Waddington CH (1942) Canalization of development and the inheritance of acquired
1215 characters. *Nature* **150**.
1216 Wang S, Sha Z, Sonstegard TS*, et al.* (2008) Quality assessment parameters for
1217 EST-derived SNPs from catfish. *BMC Genomics* **9**, 450.
1218 Werren JH, Richards S, Desjardins CA*, et al.* (2010) Functional and evolutionary
1219 insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**,
1220 343-348.
1221 Whitehead A, Crawford DL (2006a) Neutral and adaptive variation in gene
1222 expression. *Proceedings of the National Academy of Sciences of the United*
1223 *States of America* **103**, 5425-5430.
1224 Whitehead A, Crawford DL (2006b) Variation within and among species in gene
1225 expression: raw material for evolution. *Molecular Ecology* **15**, 1197-1211.
1226 Wickham H (2009) ggplot2.
1227 Yampolsky LY, Zeng E, Lopez J*, et al.* (2014) Functional genomics of acclimation
1228 and adaptation in response to thermal stress in *Daphnia*. *BMC Genomics* **15**.
1229 Yin M, Gießler S, Griebel J, Wolinska J (2014) Hybridizing *Daphnia* communities
1230 from ten neighbouring lakes: spatio-temporal dynamics, local processes, gene
1231 flow and invasiveness. *BMC Evolutionary Biology* **14**.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1232 Yin M, Wolinska J, Giessler S (2010) Clonal diversity, clonal persistence and rapid
1233 taxon replacement in natural populations of species and hybrids of the
1234 *Daphnia longispina* complex. *Molecular Ecology* **19**, 4168-4178.
1235 Zheng X, Levine D, Shen J, *et al.* (2012) A high-performance computing toolset for
1236 relatedness and principal component analysis of SNP data. *Bioinformatics* **28**,
1237 3326-3328.

1238

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

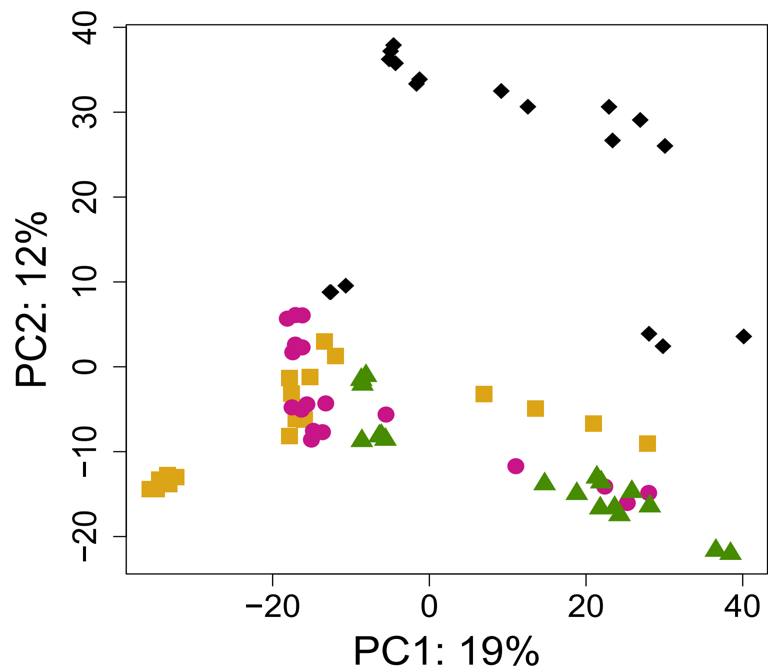
Table 1: Summary of SNP data. “NonDET” refers to transcripts that were not significantly upregulated in any of the pairwise contrasts.

bioRxiv preprint doi: https://doi.org/10.1101/374991; this version posted July 23, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

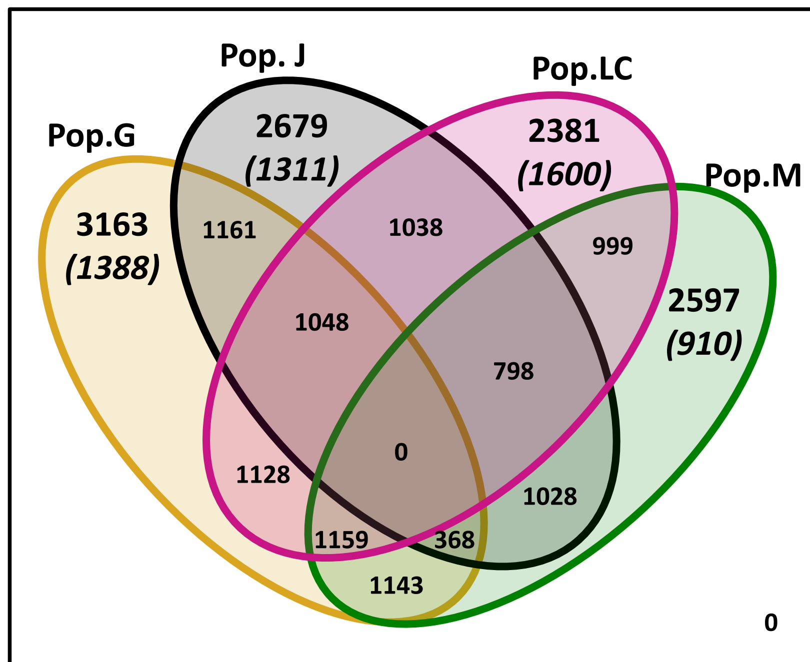
Population	D < 0	D > 0	Total
Pop.G	3,961	157	4,118
Pop.J	3,813	147	3,960
Pop.LC	4,192	87	4,279
Pop.M	3,848	164	4,012

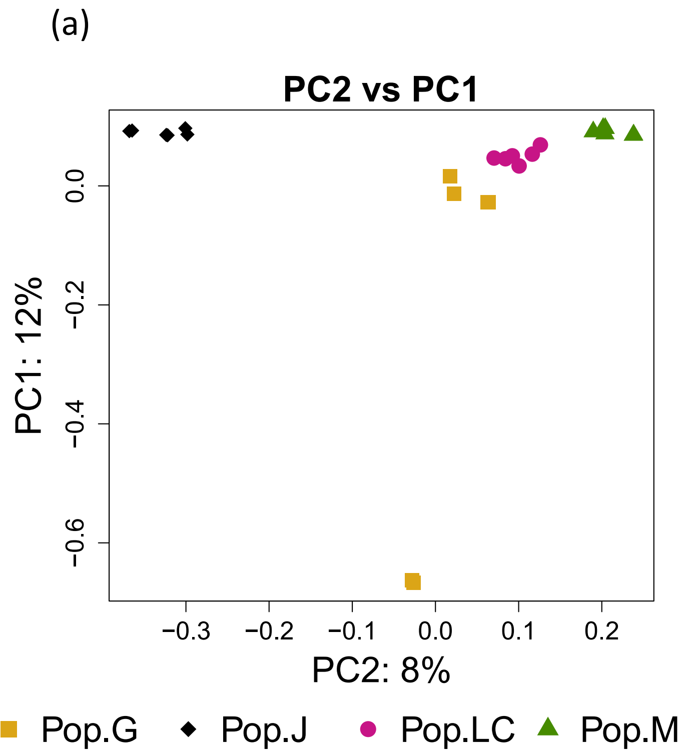
Table 2: Tajima's D test for selection. D < 0: number of transcripts with a negative Tajima's D and thus likely under purifying selection; D > 0: number of transcripts with a negative Tajima's D and thus likely under balancing selection.

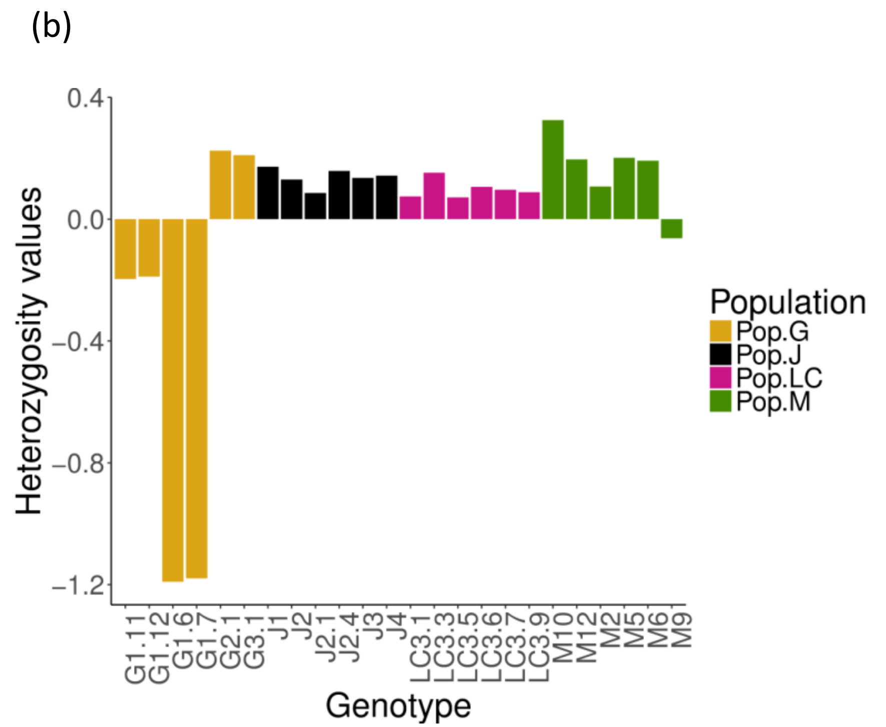
(a)



(b)



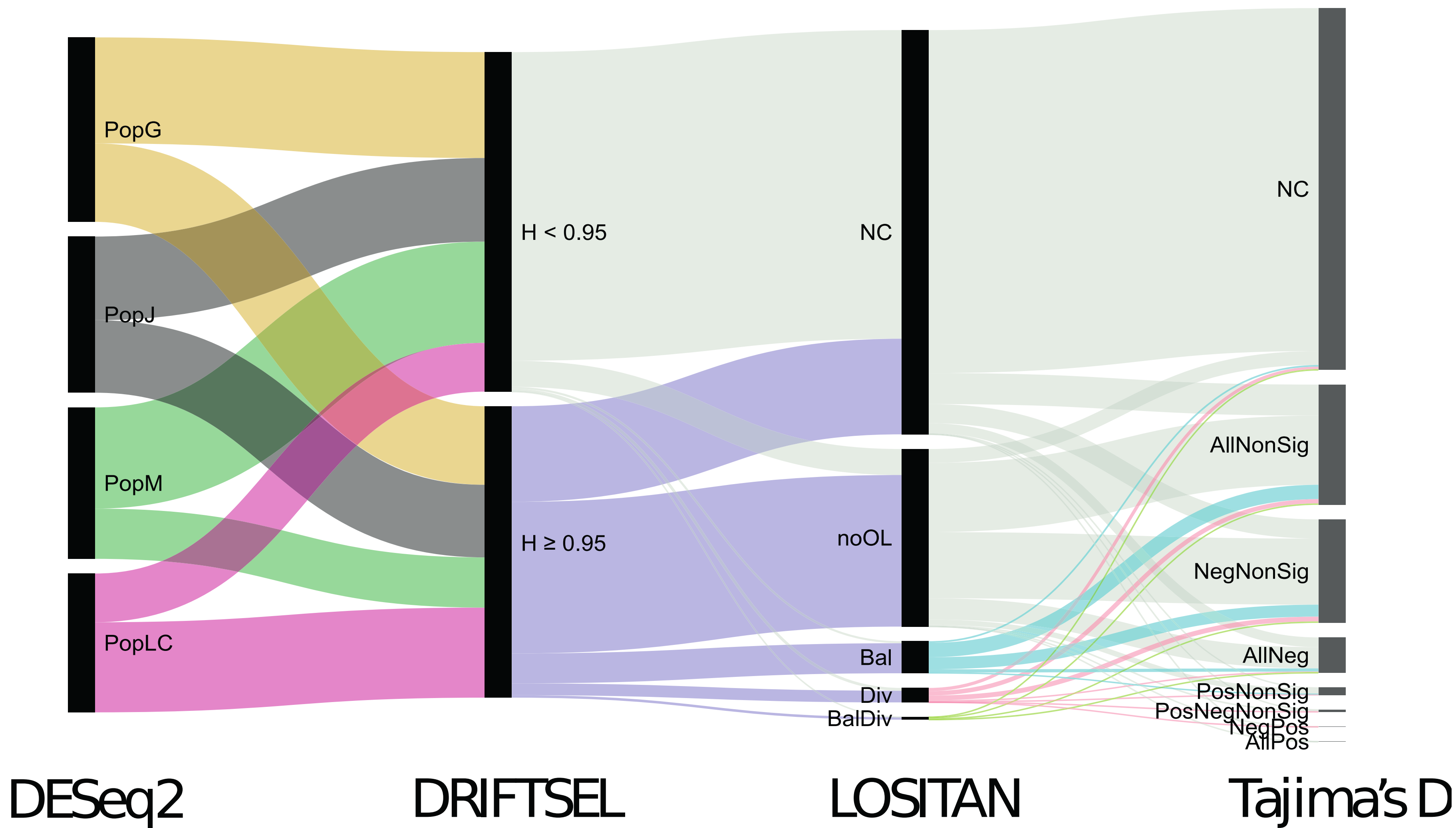

(a)
PC2 vs PC1
PC1: 12%
PC2: 8%
0.0
−0.2
−0.4
−0.6
−0.3
−0.2
−0.1
0.0
0.1
0.2
Pop.G
Pop.J
Pop.LC
Pop.M

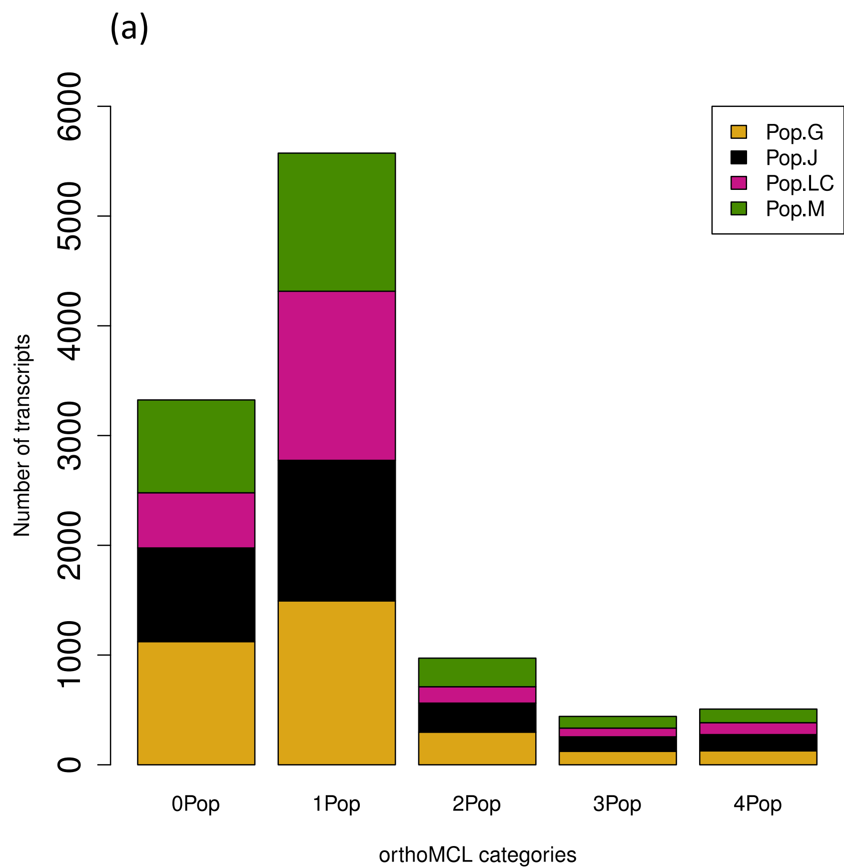
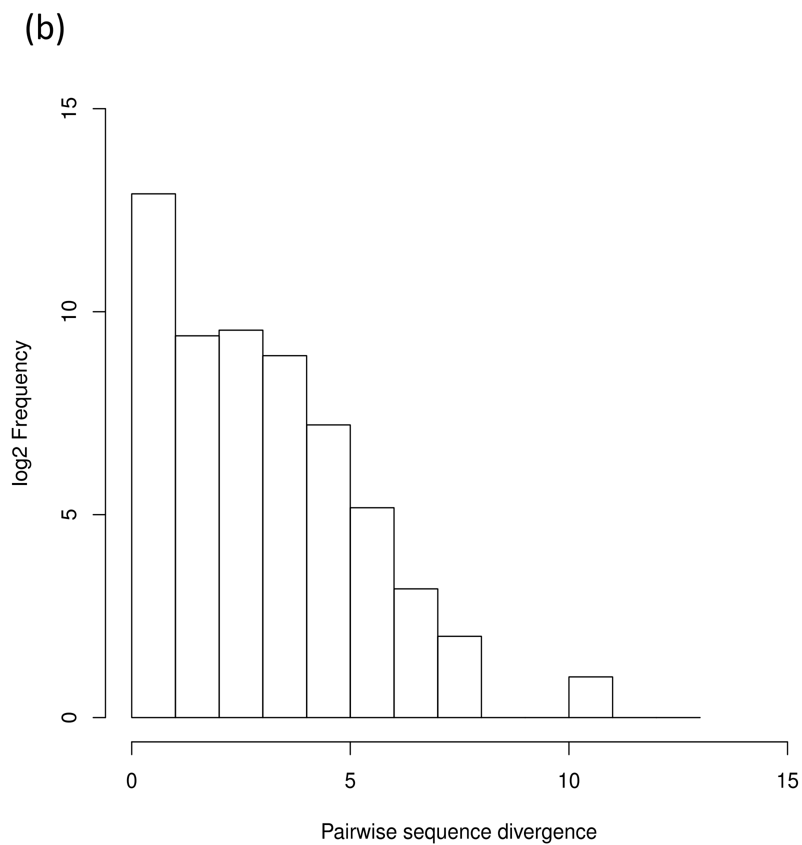

(b)
Heterozygosity values
0.4
0.0
−0.4
−0.8
−1.2
Genotype
G1.11
G1.12
G1.6
G1.7
G2.1
G3.1
J1
J2
J2.1
J2.4
J3
J4
LC3.1
LC3.3
LC3.5
LC3.6
LC3.7
LC3.9
M10
M12
M2
M5
M6
M9
Population
Pop.G
Pop.J
Pop.LC
Pop.M

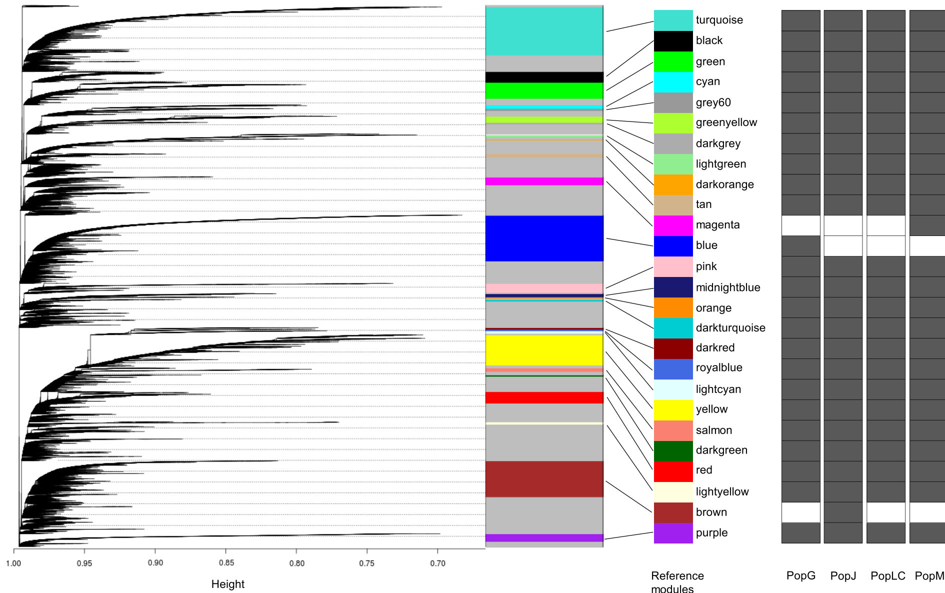
Candidates for local adaptation

Regulatory level

Sequence level




(a)
Number of transcripts
0
1000
2000
3000
4000
5000
6000
Pop.G
Pop.J
Pop.LC
Pop.M
0Pop
1Pop
2Pop
3Pop
4Pop
orthoMCL categories

(b)
log2 Frequency
0
5
10
15
0
5
10
15
Pairwise sequence divergence



Transcripts dendrogram and module colors
Module conservation (Zscores)
turquoise
black
green
cyan
grey60
greenyellow
darkgrey
lightgreen
darkorange
tan
magenta
blue
pink
midnightblue
orange
darkturquoise
darkred
royalblue
lightcyan
yellow
salmon
darkgreen
red
lightyellow
brown
purple
1.00
0.95
0.90
0.85
0.80
0.75
0.70
Height
Reference modules
PopG
PopJ
PopLC
PopM