1   **Title:** Genomic evidence for population-specific responses to coevolving parasites in a New
2   Zealand freshwater snail
3
4   **Running Title:** Lake-level responses to coevolving parasites
5

6   **Author affiliations:**
7   Laura Bankers[1], Peter Fields[2], Kyle E. McElroy[1], Jeffrey L. Boore[3], John M. Logsdon, Jr.[1],
8   Maurine Neiman[1]
9
10   [1]Department of Biology, University of Iowa, 143 Biology Building, Iowa City, IA 52242, USA
11   [2]Zoologisches Institut, Universität Basel, Vesalgasse 1, Basel, CH-4051, Switzerland
12   [3]Department of Integrative Biology, University of California, 3040 Valley Life Sciences
13   Building, Berkeley, CA 94720, USA
14
15
16
17   **Corresponding author***:
18   Laura Bankers
19   Department of Biology
20   University of Iowa
21   143 Biology Building
22   Iowa City, IA 52242, USA
23   laura-rice@uiowa.edu
24
25

26   **Keywords:** host-parasite interactions, coevolution, transcriptomics, molecular evolution,
27   molluscs

1

Abstract

28

29   Reciprocal coevolving interactions between hosts and parasites are a primary source of strong

30   selection that can promote rapid and often population- or genotype-specific evolutionary change.

31   These host-parasite interactions are also a major source of disease. Despite their importance, very

32   little is known about the genomic basis of coevolving host-parasite interactions in natural

33   populations, especially in animals. Here, we use gene expression and sequence evolution

34   approaches to take critical steps towards characterizing the genomic basis of interactions between

35   the freshwater snail *Potamopyrgus antipodarum* and its coevolving sterilizing trematode parasite,

36   *Microphallus* sp., a textbook example of natural coevolution. We found that *Microphallus*-

37   infected *P. antipodarum* exhibit systematic downregulation of genes relative to uninfected *P.*

38   *antipodarum*. The specific genes involved in parasite response differ markedly across lakes,

39   consistent with a scenario where population-level coevolution is leading to population-specific

40   host-parasite interactions and evolutionary trajectories. We also used an F_{ST}-based approach to

41   identify a set of loci that represent promising candidates for targets of parasite-mediated selection

42   across lakes as well as within each lake population. These results constitute the first genomic

43   evidence for population-specific responses to coevolving infection in the *P. antipodarum*-

44   *Microphallus* interaction and provide new insights into the genomic basis of coevolutionary

45   interactions in nature.

46

Introduction

47

48   Host-parasite interactions constitute a primary source of natural selection and provide a powerful

49   means of evaluating the evolutionary response to strong selection (Hamilton 1980; Abbate *et al.*

50   2015). Reciprocal antagonistic selection for greater host resistance and parasite infectivity can

51   lead to antagonistic coevolution between host and parasite (Ehrlich & Raven 1964), which can in

52   turn maintain high genetic diversity (Laine 2009; Bérénos *et al.* 2011; King *et al.* 2011; van

53   Houte *et al.* 2016) and drive rapid evolutionary diversification (Laine 2009; Paterson *et al.* 2010;

54   Masri *et al.* 2015).

55   While a great deal is known about the genomic underpinnings of plant-pathogen

56   interactions (reviewed in Kover & Caicedo 2001), much less is known about animal systems, and

57   particularly in natural populations (Routtu & Ebert 2015). Host-parasite interactions often have a

58   genetic basis (*e.g.,* Paterson *et al.* 2010; Luijckx *et al.* 2013; Masri *et al.* 2015; van Houte *et al.*

59   2016), are linked to rapid molecular evolution (*e.g.,* Paterson *et al.* 2010; Tennessen *et al.* 2015),

60   and are associated with gene expression changes (*e.g.,* Barribeau *et al.* 2014; de Bekker *et al.*

61   2015; McTaggart *et al.* 2015). Because most studies providing important insights into the genetic

62   and genomic basis of host-parasite coevolutionary interactions have been limited to the laboratory

63   (*e.g.,* Roger *et al.* 2008; Arican-Goktas *et al.* 2014; Barribeau *et al.* 2014; but also see

64   Decaestecker *et al.* 2007; Routtu & Ebert 2015), we know very little about the evolutionary

65   genomics of host-parasite interactions in natural populations (especially animals) and whether

66   results from laboratory studies are recapitulated in nature (Routtu & Ebert 2015). As such,

67   elucidating the processes of coevolution in nature is a requirement both to characterize a primary

68   driver of selection and to formulate targeted strategies to fight natural parasite populations (*e.g.,*

69   Gleichsner *et al.* 2015; Neafsey *et al.* 2015). Characterization of parasite-mediated changes in

70   host gene expression provides a powerful means of deciphering the genomic basis of natural

71   coevolutionary interactions because expression changes are often genotype or population-specific

72   (*e.g.,* Barribeau *et al.* 2014), can affect host response to parasites (*e.g.,* Brunner *et al.* 2013), and

73   can underlie variation in resistance and susceptibility to infection (*e.g.,* Stutz *et al.* 2015).

74   The New Zealand freshwater snail *Potamopyrgus antipodarum* and its sterilizing

75   coevolving trematode parasite *Microphallus* sp. ("*Microphallus*") provide an especially

76   compelling context for characterizing the genomic basis of coevolution in natural populations.

77   First, selection imposed by both host and parasite is extremely strong: infected snails are

78   completely sterilized (Winterbourn 1973), and *Microphallus* that cannot infect the first *P.*

79   *antipodarum* by which they are ingested will die (King *et al.* 2011). Second, different *P.*

80   *antipodarum* populations experience consistently high *vs.* consistently low *Microphallus*

81   infection frequencies (Lively & Jokela 2002; King & Lively 2009; Vergara *et al.* 2013),

82   indicating that infection frequency is a major determinant of the strength of parasite-mediated

83   selection and, thus, the rate and mode of coevolution within each population. Third, *Microphallus*

84   is locally adapted to *P. antipodarum* at both within- and among-lake scales (Lively *et al.* 2004;

85   King *et al.* 2009), demonstrating the fine spatial scale of coevolution in this system. Fourth, the

86   existence of many naturally replicated and separately evolving *P. antipodarum-Microphallus*

87   interactions means that each population can be treated as a separate evolutionary experiment into

88   the consequences of antagonistic coevolution (Dybdahl & Lively 1996). Finally, rapid

89   coevolution of *P. antipodarum* and *Microphallus* has been documented in both a natural

90   population (Jokela *et al.* 2009) and in an experimental coevolution study (Koskella & Lively

91   2009).

92   Multiple studies suggest that coevolutionary interactions between *P. antipodarum* and

93   *Microphallus* fit a "matching alleles" infection genetics model, whereby there are no universally

94   infective parasites or resistant hosts (Dybdahl & Krist 2004; Lively *et al.* 2004; Osnas & Lively

95   2005; King *et al.* 2009). Rather, *P. antipodarum* susceptibility depends on whether the genotype

96   of the *Microphallus* individual matches the *P. antipodarum* individual at loci involved in

97   resistance. Similar matching-allele mechanisms are thought to operate in other snail-trematode

98   systems such as the laboratory model *Biomphalaria glabrata-Schistosoma mansoni* (reviewed in

99   Adema & Loker 2015). Recent studies suggest that both allelic identity (Roger *et al.* 2008) and

100   gene expression (Arican-Goktas 2014) are likely involved in mediating *B. glabrata* susceptibility

101   to *S. mansoni* infection, emphasizing the potentially central role for gene expression in

102   determining outcomes of host-parasite interactions in this and other snail-trematode systems

103   (Hotez 2013; Jurberg & Brindley 2015).

104   Here, we take critical steps towards illuminating the genomic basis of coevolution in this

105   textbook example of antagonistic coevolution (*e.g.*, Zimmer & Emlen 2013; Herron & Freeman

106   2014; Bergstrom & Dugatkin 2016) by using RNA-Seq to perform gene expression and F_{ST}

107   analyses for three replicated samples of *Microphallus*-infected *vs.* uninfected *P. antipodarum*

108   from each of three different lake populations. In light of existing ecological evidence for

109   population structure and local adaptation in this system, we expected to observe divergence

110   between populations at both the gene expression and genetic level, such that population of origin

111   as well as infection status would influence patterns of gene expression and relative levels of

112   genetic differentiation. Such divergence at genetic and gene expression levels should be

113   especially notable with respect to genes likely to be involved in the coevolutionary response to

114   *Microphallus* infection.

115

116   **Materials and Methods**

117   **Sample collection, dissection, and sample selection.** Adult *P. antipodarum* were collected from

118   shallow water (depth < 1m) habitats along the lake margins of three New Zealand lakes

119   (Alexandrina, Kaniere, Selfe) known to contain relatively high *Microphallus* infection

120   frequencies (~10-20%, Vergara *et al.* 2013; Table S1). Following transfer to the University of

121   Iowa, we housed snails in 15 L tanks at 16°C with a 16:8 hour light:dark cycle and supplied *ad*

122   *libitum* dried *Spirulina* (*e.g.*, Zachar & Neiman 2013) for one week to allow snails to recover

123   from international travel prior to dissection and RNA isolation.

124   Because *P. antipodarum* is polymorphic for reproductive mode and ploidy level (Lively

125   1987; Neiman *et al.* 2011), we confined our RNA-Seq analyses to diploid adult non-brooding

126   (non-reproductively active) females. These criteria ensured that we limited extraneous biological

127   processes that may conflate differences in gene expression related to interactions between *P.*

128   *antipodarum* and *Microphallus*. We dissected each snail to determine sex (male *vs.* female),

129   *Microphallus* infection status (infected *vs.* uninfected), and reproductive status (brooding *vs.* non-

130   brooding). We used the infection data to establish infection frequency for each lake sample

131   (Table S1). Because *Microphallus* infection fills the body cavity of *P. antipodarum*, we confined

132   analyses to head tissue. While the use of head tissue necessarily prevents the analysis of genes

133   that are solely expressed in body tissue, this approach is the only way to ensure that comparable

134   tissue types are isolated from both infected and uninfected snails.

135   All infected snails had stage five *Microphallus* infections

136   (http://www.indiana.edu/~curtweb/trematodes/DATA_KEY.HTM). We first separated the head

137   tissue of dissected snails, which does not contain *Microphallus* metacercariae, from body tissue,

138   and then split the dissected head tissue into two halves. We immediately submerged one half of

139   the head tissue in RNA*later*® Solution (Life Technologies Corporation) and stored the tissue at

140   4°C for 24 hours followed by storage at -80°C (according to manufacturer protocol) until RNA

141   extraction. The other head half was immediately snap-frozen in liquid nitrogen and stored at -

142   80°C for flow cytometric determination of ploidy level following Krist *et al.* (2014). For each

143   RNA-sequencing replicate, we pooled head tissue from seven snails to obtain a sufficient amount

144   of tissue for RNA extraction and sequencing. We obtained three biological replicates of parasite-

145   infected and uninfected *P. antipodarum* from each of the three lakes, for a total of 18 replicates

146   and 126 snails.

147

148   **RNA sequencing and *de novo* reference transcriptome assembly and annotation.** We

149   extracted RNA following the Invitrogen TRIzol protocol (Chomczynski 1993). RNA quantity and

150   quality were assessed with a Bio-Rad Experion Automated Electrophoresis Station, following

151   manufacturer protocol for the Experion RNA analysis kit with an RQI $\geq$ 8 and a minimum of 2μg

152   of total RNA. RNA shearing and cDNA library preparation were completed following the

153   Illumina Truseq LS protocol (Illumina, San Diego, CA, 2012). Following library preparation, we

154   used an Illumina HiSeq 2000 for 2x100 bp paired-end RNA sequencing. Each RNA sample was

155   given a unique indexed adapter sequence (Illumina, San Diego, CA, 2012) and was then

156   separated into two halves, allowing us to sequence each sample twice in two different lanes in

157   order to eliminate bias due to sequencing lane (two technical replicates per sample). We obtained

158   a mean read length (SD) of 100.7 bp (0.96 bp) and a mean number of paired-end reads/replicated

159   sample/lane (SD) of 10070420 (1474539.10), for ~20000000 paired-end reads/replicate. Next, we

160   used FASTX Toolkit (Gordon & Hannon 2010) and FastQC (Andrews 2010) to trim adapters,

161   assess sequencing quality, and filter out low-quality reads from the raw RNA-Seq data (see Table

162   S2 for specific parameters). Quality filtering resulted in a mean sequence quality (Phred) score of

163   35.95, and a mean number of reads/replicate/lane (SD) of 9871440.4 (1468645.04), for a total of

164   ~18000000 paired-end reads/replicate, about twice the coverage required to quantify expression

165   differences across a wide range of expression levels (Wolf 2013).

7

166 We used Trinity v. 2.0.4 (Gabherr *et al.* 2011; Haas *et al.* 2013) to generate an initial *de*

167 *novo* assembly using all of the combined filtered RNA-Seq data. First, we performed the

168 recommended Trinity *in silico* normalization for each replicate in order to reduce the amount of

169 memory required for the assembly process while maintaining a representative read set (following

170 Gabherr *et al.* 2011; Haas *et al.* 2013). We then used Trinity to assemble the normalized reads

171 (Haas *et al.* 2013), generating a transcriptome assembly with 462736 transcripts (Table S2). We

172 annotated protein-coding regions in the transcriptome in order to identify long ORFs of putative

173 genes that could be missed by homology searches and to filter out miscalled isoforms by using

174 the Trinity plugin TransDecoder (Haas *et al.* 2013) (Table S2). This step also lessens the

175 influence of sequencing errors, misassembled transcripts, chimeric sequences, and other common

176 assembly issues (De Wit *et al.* 2015). Next, we used hierarchical clustering based on sequence

177 identity to further reduce redundancy in the transcriptome assembly, as implemented in CD-HIT-

178 EST (Huang *et al.* 2010) (Table S2). Finally, we identified and eliminated potential contaminant

179 transcripts using the 'blast_taxonomy_report.pl' script from the Blobology pipeline (Kumar *et al.*

180 2013), blastx (Camacho *et al.* 2009) and a custom python script (available at:

181 github.com/jsharbrough/grabContigs). We used Blobology to filter out transcripts with non-

182 metazoan top blast hits, as well as Platyhelminthes (representing potential *Microphallus*

183 contamination). These last filtering steps provided a final reference transcriptome assembly with

184 62862 transcripts.

185 We annotated the reference transcriptome by using blastx (Camacho *et al.* 2009) with an

186 E-value cutoff of 1e-5. We then imported these blastx results into Blast2GO to assign gene

187 ontology ("GO") terms to blastx-annotated transcripts (Conesa *et al.* 2005). We obtained blastx

188 annotations for 10171 transcripts and both blastx and GO annotations for 15797 transcripts.

189   Nearly 75% of our transcriptome did not receive GO annotations, meaning that the functions of

190   the majority of genes in our transcriptome cannot yet be determined. This result is unsurprising in

191   light of a distinct deficit of mollusc genome sequence data available to aid in annotation (Kocot *et*

192   *al*. 2015); though molluscs are the most species-rich animal phylum after arthropods (Smith *et al.*

193   2011), only ten (~1.4%) of the 700 sequenced animal genomes available on NCBI as of

194   September 2016 are from molluscs.

195

196   **RNA-Seq gene expression analyses, functional enrichment, and F_{ST} outlier analyses**. We

197   used our *de novo* transcriptome as our reference for all gene expression analyses. First, we used

198   Tophat2 (Trapnell *et al.* 2013) to map filtered RNA-Seq reads to our *de novo* reference

199   transcriptome. Next, we assembled mapped reads into transcripts and estimated transcript

200   abundance with Cufflinks (Trapnell *et al.* 2012). We merged Cufflinks GTF and Tophat bam files

201   for use in CuffDiff with Cuffmerge (Trapnell *et al.* 2012). Finally, we identified and quantified

202   significant changes in gene expression using an FDR of 5% and Benjamini-Hochberg multiple

203   test correction, as implemented in CuffDiff (Trapnell *et al.* 2012). We required that fragments per

204   kilobase per million reads mapped (FPKM) exceeded zero to ensure that we were only comparing

205   expression patterns among genes transcribed in all replicates (Table S2). We visualized the gene

206   expression results with heatmaps, MDS plots, and dendrograms generated with cummeRbund

207   (Goff *et al.* 2013), and used Blast2GO to compare the functions of differentially expressed genes

208   and quantify the number of transcripts annotated with each GO term. We performed functional

209   enrichment analyses and Fisher's Exact Tests as implemented in Blast2GO to identify

210   significantly overrepresented functional groups among differentially expressed genes (Table S2).

211   We used F_{ST} outlier analyses to identify those genes with especially high levels of genetic

212   differentiation between infected *vs*. uninfected snails; these genes provide a set of candidates for

213   *Microphallus*-mediated selection. We filtered our reference transcriptome to include only

214   transcripts with fragments per kilobase per million reads mapped (FPKM) > 0 in all 18 replicates

215   (30685 transcripts) to ensure each replicate was represented for all loci in the F_{ST} comparisons.

216   We used Tophat2 (Trapnell *et al.* 2013) with the same parameters as in the gene expression

217   analyses to map RNA-Seq reads to the filtered transcriptome. We then used Picard Tools to

218   prepare mapped reads for variant discovery (http://picard.sourceforge.net), applying the

219   AddOrReplaceReadGroups script to add read groups to each mapped bam file. We merged

220   technical replicate bam files with the MergeSamFiles script, resulting in a single bam file for each

221   of the 18 replicates. Finally, we used the MarkDuplicates script to identify duplicate reads,

222   removing reads that mapped to more than one location in the transcriptome. Next, we used

223   Samtools mpileup (Li *et al.* 2012) to call SNPs from processed bam files.

224   Analyses of levels of nucleotide variation from pooled RNA-sequencing data with

225   programs like Popoolation2 (Kofler *et al.* 2011), which was developed to perform variant calling

226   and calculate F_{ST} from pooled sequencing data, can effectively identity SNPs and candidate focal

227   genes of selection and perform F_{ST}-based measures of genetic differentiation (*e.g.*, Fischer *et al.*

228   2013; Konczal *et al.* 2014). We used Popoolation2 (Kofler *et al.* 2011) to calculate F_{ST} per site.

229   The Popoolation2 pipeline begins by using the mpileup2sync.pl script to generate synchronized

230   mpileup files that were filtered for base quality (Q20). We then applied the fst-sliding.pl script to

231   calculate F_{ST} per site. In all cases, we required a minimum coverage of ten, a maximum coverage

232   of 200 (to reduce memory usage and biases introduced by inter-locus variation in gene

233   expression), and that each SNP was sequenced at least four times (*e.g.*, Fischer *et al.* 2013;

234   Konczal *et al.* 2014). We otherwise used default parameters. Finally, we filtered out SNPs with

235   minor allele frequencies less than 5%. We chose these parameters to account for sequencing

236   errors and to reduce the likelihood of obtaining false positives by ensuring that each SNP was

237   sequenced multiple times, at the risk of not detecting rare alleles (*e.g.,* Konczal *et al.* 2014). We

238   used IBM SPSS Statistics v. 23 to perform outlier analyses and identify outlier SNPs between

239   infected and uninfected snails. We also identified the expression patterns of each transcript

240   containing F_{ST} outlier SNPs based on the expression analyses detailed above and evaluated

241   whether transcripts containing F_{ST} outliers between infected and uninfected snails contained

242   outliers in multiple lake populations *vs.* only a single lake population. Finally, we compared

243   relative levels of genetic differentiation (mean per-SNP F_{ST} across all sites for which F_{ST} could be

244   estimated) between infected and uninfected snails within each lake and mean per-SNP F_{ST}

245   between lakes using Bonferroni-corrected Welch's t-tests as implemented within IBM SPSS

246   Statistics v. 23. The putative functions of transcripts representing F_{ST} outliers were assessed using

247   Blast2GO.

248   We used three complementary analytical approaches to evaluate the effect of infection

249   and population on gene expression and genetic differentiation (Fig. S1). First, for the "inclusive

250   analysis," we identified differentially expressed genes likely to be broadly important for *P.*

251   *antipodarum* response to *Microphallus* infection. We began by quantifying expression for each

252   transcript for all pooled replicates under both conditions (infected and uninfected) to determine

253   expression differences between all infected *vs.* all uninfected snails (Fig. S1a). We then compared

254   the annotated functions of differentially expressed genes and performed functional enrichment

255   analyses to determine the types of genes broadly important for infection response. Next, we

256   calculated mean per-SNP F_{ST} and performed F_{ST} outlier analyses to measure relative levels of

257   genetic differentiation within and between infected and uninfected snails and identified functional

258   annotations and expression patterns of each transcript containing one or more F_{ST} outlier SNPs.

259   Second, for the "within-lake analysis", we compared gene expression patterns between

260   infected *vs.* uninfected snails from each lake to characterize local (lake level) gene expression

261   responses to *Microphallus* infection. This analysis included pairwise comparisons between

262   infected and uninfected snails from within each lake population (Fig. S1b). We then compared

263   the functions of differentially expressed genes between infected and uninfected snails within each

264   lake. We also calculated mean per-SNP F_{ST} and performed F_{ST} outlier analyses between infected

265   and uninfected snails within each lake. Finally, we determined the functional annotation and

266   expression pattern of each outlier-containing transcript.

267   Third, in the "across-lake analysis", we compared replicates by lake and infection status.

268   Here, our goal was to use patterns of gene expression within and across lakes to evaluate

269   evidence for local adaptation and identify genes likely to be locally important for response to

270   *Microphallus* infection. We conducted every possible pairwise comparison of gene expression

271   between all replicates from all three lakes (Fig. S1c), allowing us to differentiate between genes

272   that were differentially expressed in infected *vs.* uninfected replicates across the three lakes (*i.e.*,

273   evidence for population-specific infection responses) and genes expressed differently across lakes

274   regardless of infection status (*i.e.*, expression difference due to lake of origin rather than infection

275   status). Annotation and functional enrichment analyses allowed us to identify putative functions

276   of genes that were significantly differentially expressed between infected and uninfected snails in

277   more than one lake population *vs.* within a single lake. We visualized these comparisons with

278   Euler diagrams generated with eulerAPE v3 (Micallef & Rodgers 2014). Finally, we evaluated

279   whether transcripts containing F_{ST} outliers between infected and uninfected snails contained

280   outliers in multiple lake populations *vs.* only a single lake population.

281

282   **Results**

283   **Inclusive analysis: Greater downregulation of transcripts in infected snails.** We identified

284   1408 significantly differentially expressed transcripts (FDR: 5%, Benjamini-Hochberg) between

285   *Microphallus*-infected and uninfected snails (Figs. 1a, S1a). Here, we define "upregulated" and

286   "downregulated" as significantly higher and lower, respectively, expression in infected snails

287   relative to uninfected counterparts. A significantly higher proportion of these 1408 transcripts

288   were downregulated (1057, ~75%) *vs.* upregulated (351, ~25%) (Fisher's Exact Test: $p < 0.0001$;

289   Table 1, Fig. 1a), indicating that infected *P. antipodarum* experience systematic reduction in gene

290   expression.

291   We obtained gene ontology (GO) annotations for 447 of these 1408 transcripts, 216 of

292   which were significantly upregulated and 231 significantly downregulated in infected snails.

293   Sixteen of these upregulated genes were annotated as involved in immune system processes (Fig.

294   S2), making these loci our strongest candidates for direct involvement in response to

295   *Microphallus* infection. The ten genes with putative brain/behaviour functions that were

296   upregulated in infected snails (Fig. S2) are particularly interesting candidates for future study in

297   light of the well-characterized influence of *Microphallus* infection on *P. antipodarum* behaviour

298   (*e.g.*, Levri 1999). Functional enrichment revealed that in infected snails, antigen processing and

299   presentation were overrepresented among upregulated genes. Processes related to transcription

300   and translation were overrepresented among downregulated genes, consistent with our

301   observation that infected snails had significantly more downregulated genes (Table S3).

13

302   Our inclusive F_{ST} analysis, comparing relative levels of genetic differentiation between

303   infected and uninfected snails, revealed a mean per-SNP F_{ST} (SD) of 0.0367 (0.0504) and

304   identified 70 outlier SNPs (mean F_{ST} of outliers (SD) of 0.3849 (0.2038)) from 59 transcripts. Of

305   these 59 transcripts, ten were downregulated and three were upregulated in infected snails,

306   meaning the majority of F_{ST} outlier-containing genes were not significantly differentially

307   expressed (Table 1). We obtained functional annotations for 16 of the 59 transcripts, four of

308   which are likely involved in immune system processes (one transcript) and response to stimuli

309   (three transcripts) (Tables S4, S5).

310

311   **Within-lake analysis: Lake-specific responses to *Microphallus* infection.** Similar to the

312   inclusive analysis, we observed a significantly greater proportion of downregulated *vs*.

313   upregulated transcripts in infected *P. antipodarum* in two of the three lakes (Table 1, Figs. 1b,

314   S1b). A total of 1539 transcripts were significantly differentially expressed between infected and

315   uninfected snails in at least one of the three pairwise tests (Fig. 1b).

316   Functional enrichment for the within-lake analysis revealed that various metabolic

317   processes are overrepresented among upregulated genes in infected snails from Alexandrina

318   (Table S3). We did not detect significant functional enrichment for genes upregulated in infected

319   snails from lakes Selfe or Kaniere. Upregulated genes in uninfected snails from Alexandrina and

320   Selfe are enriched for GO terms related to transcription and translation. Upregulated genes in

321   uninfected snails from Kaniere are enriched for pantothenate metabolism (Table S3).

322   The mean per-SNP F_{ST} (SD) between infected and uninfected snails within each lake was

323   0.0485 (0.0504), 0.0479 (0.0623), and 0.0484 (0.0617) for Alexandrina, Kaniere, and Selfe,

324   respectively (Fig. 2), providing evidence for population-specific levels of genetic differentiation

325   between infected and uninfected snails. We identified 42, 50, and 47 F_{ST} outlier SNPs between

326   infected and uninfected snails (mean F_{ST} of outliers (SD) of 0.3153 (0.1346), 0.3612 (0.1536),

327   and 0.3352 (0.1442) from Alexandrina, Kaniere, and Selfe, respectively). Similar to the inclusive

328   analysis, the vast majority of outlier-containing transcripts (93.5%) were not significantly

329   differentially expressed (Table 1).

330

331   **Across-lake analysis: Lake of origin strongly influences gene expression and relative levels**

332   **of genetic differentiation**. We observed 6228 significantly differentially expressed transcripts

333   within and across lakes (Figs. 3, S1c). ~75% (4689) of these 6228 transcripts were significantly

334   differentially expressed across the three lakes regardless of infection status, indicating that lake of

335   origin has a markedly stronger influence on gene expression than infection status (Figs. 1b, 3, 4,

336   S3). Similarly, when replicates are clustered based on expression profile, they group first by lake

337   of origin, followed by infection status. This result indicates both that population of origin is a key

338   determinant of gene expression in *P. antipodarum* (Figs. 3, S3) and that infection results in

339   predictable changes in expression (Fig. 3) that are still detectable on this background of

340   population divergence. These findings are consistent with and extend to the gene expression level

341   evidence for population-specific phenotypes (Holomuzki & Biggs 2006; Krist *et al.* 2014) and

342   marked population genetic structure (Neiman & Lively 2004; Neiman *et al.* 2011; Paczesniak *et*

343   *al.* 2013) in *P. antipodarum*. These results also demonstrate that both population of origin and

344   infection have marked consistent and systematic consequences for gene expression in this

345   species.

346   We used the total number of transcripts identified as significantly differentially expressed

347   between infected and uninfected snails (1539) in at least one of the three within-lake analyses

348   (Fig. 1b) to compare the number of significantly up or downregulated transcripts in multiple *vs*.

349   single populations, allowing us to determine whether snails from different lakes have similar gene

350   expression responses to *Microphallus* infection. This comparison revealed that nearly all of the

351   differentially expressed transcripts from the within-lake analysis (1447 transcripts, 94%) are only

352   significantly differentially expressed in a single lake. Only 6% (92) of the differentially expressed

353   transcripts showed significant differential expression in more than one population (Fig. 4). In

354   summary, the vast majority of differentially expressed transcripts between infected and

355   uninfected snails show significant up or downregulation in only one population, suggesting a

356   distinct local (lake specific) gene expression response to parasite infection.

357   Next, we identified the types of genes that were significantly differentially expressed in

358   single *vs*. multiple populations to parse out general *vs*. lake-specific infection responses. Of the 92

359   genes that were significantly differentially expressed in more than one lake, 22 genes were

360   upregulated in infected snails (Table S6). We obtained GO annotations for seven of these genes,

361   which had functions related to immune response, nervous system function, and metabolism. Ten

362   of the 70 genes found to be significantly downregulated in infected snails in more than one lake

363   received GO annotations, which included immune function, response to stimulus, and

364   transcription/translation (Table S6). 396 of the 1447 genes that were significantly differentially

365   expressed in only one lake received GO annotations (Fig. S2). Even though the particular genes

366   experiencing differential expression differ on a lake-by-lake basis, these different genes often

367   belong to similar GO categories (*e.g.*, immune system processes, response to stimulus, and

368   behaviour).

369   We also evaluated how many transcripts contained F_{ST} outlier SNPs in single *vs*. multiple

370   populations. Similar to our gene expression analyses, we found that very few transcripts (six

371 transcripts, ~5%) containing F_{ST} outlier SNPs between infected and uninfected snails in one

372 population also contained outlier SNPs in another population (three transcripts between

373 Alexandrina and Kaniere, two transcripts between Alexandrina and Selfe, and one transcript

374 between Kaniere and Selfe). This result indicates that the focal loci of *Microphallus*-mediated

375 selection are likely to often be population specific. Of the transcripts that contained outlier SNPs,

376 12, 19, and 16 transcripts were annotated for Alexandrina, Kaniere, and Selfe, respectively. Two

377 of these transcripts were annotated as relevant to immune system processes, 13 transcripts were

378 annotated as involved in response to stimuli, and six transcripts were annotated with functions

379 related to neurological/behavioural processes (Tables S4, S5).

380   Finally, we compared mean per-SNP F_{ST} across lakes as a whole, for infected replicates

381 only, and for uninfected replicates only. We found that F_{ST} between infected and uninfected

382 snails within each lake was significantly lower than the mean F_{ST} for any across-lake comparison

383 (Fig. 2), indicating relatively greater levels of genetic differentiation among lake populations than

384 between infected and uninfected snails within a population. This result is consistent with and

385 extends to the genome level the outcome of marker-based studies (*e.g.,* Paczesniak *et al.* 2013)

386 that have documented strong across-lake genetic structure in *P. antipodarum*.

387

388 **Discussion**

389   We used gene expression and F_{ST} analyses to shed light on the genomic underpinnings of

390 coevolution in a natural context. Most importantly, we found that population of origin is an

391 important factor influencing gene expression, and that infection of *P. antipodarum* by its

392 coevolving trematode parasite *Microphallus* elicits a marked, systematic, and population-specific

393 gene expression response. These results have important implications for our understanding of the

17

394   dynamics of coevolution in the *P. antipodarum-Microphallus* system, setting the stage for future

395   research targeted at in-depth characterization of these population-level effects on the candidate

396   loci that we identified. Our F_{ST} analyses allowed us to evaluate relative levels of genetic

397   differentiation within and among lakes and also revealed promising candidate genes for the focus

398   of *Microphallus*-mediated selection. These findings provide a qualitative advance by extending

399   evidence for local adaptation and coevolution in this textbook coevolutionary interaction to the

400   genomic level, demonstrating distinct local genetic and gene expression responses by *P.*

401   *antipodarum* to *Microphallus* infection. Together, these results illuminate the unique and

402   important insights that can come from sampling multiple natural populations of interacting hosts

403   and parasites.

404

405   **Systematic downregulation and response to parasite infection**. Our inclusive analysis revealed

406   that the majority of transcripts that are significantly differentially expressed in infected snails are

407   downregulated relative to uninfected snails. This pattern could reflect several non-mutually

408   exclusive phenomena, ranging from tissue/organ destruction and/or overall poor condition of

409   infected snails (*e.g.,* Barribeau *et al.* 2014) to reallocation of resources to genes needed for

410   defence against and response to *Microphallus* infection (*e.g.,* Ederli *et al.* 2015) to suppression

411   and/or parasitic manipulation of *P. antipodarum* gene expression by *Microphallus* as a means of

412   evading host immune and defence systems (*e.g.,* Levri & Lively 1996; Barribeau *et al.* 2014;

413   Soper *et al.* 2014; de Bekker *et al.* 2015). Regardless of the specific mechanism(s) involved, the

414   transcripts that are differentially expressed between infected and uninfected snails in the inclusive

415   analysis represent a set of genes that are most likely to be of general importance to the *P.*

416   *antipodarum-Microphallus* interaction. Our inclusive F_{ST} analyses yielded additional candidate

18

417   genes for the focus of parasite-mediated selection; the 70 SNPs from 59 genes that show

418   relatively high levels of genetic differentiation provide a particularly strong set of candidate loci

419   for *Microphallus* response. The presence of genes annotated with immune/stress responses and

420   with neurological and behaviour-related functions strengthens this conclusion.

421

422   **Genomic evidence for population-level responses to infection**. Our results demonstrate that *P.*

423   *antipodarum* response to *Microphallus* infection is largely population specific. For example, in

424   Alexandrina and Kaniere, the majority of significantly differentially expressed transcripts are

425   downregulated in infected relative to uninfected snails (73% for Alexandrina and 66% for

426   Kaniere). By contrast, the proportion of significantly upregulated and downregulated transcripts

427   is much more similar in snails from Selfe (52% *vs*. 58%, respectively). This pattern is interesting

428   in light of data from Vergara *et al.* (2013) suggesting that lake Selfe *P. antipodarum* have

429   historically experienced lower frequencies of *Microphallus* infection (3.5x lower than

430   Alexandrina, 1.4x lower than Kaniere; Table S1) (Vergara *et al.* 2013), suggesting that these

431   populations may have experienced different historical intensities of *Microphallus*-mediated

432   selection.

433   The vast majority (~87%) of differentially expressed transcripts were only significantly

434   differentially expressed in a single lake population, and ~75% of significantly differentially

435   expressed transcripts were expressed differently across lakes regardless of infection status,

436   highlighting the important influence of population of origin on gene expression. We also found

437   significantly lower levels of relative genetic differentiation between snails within *vs.* across lakes

438   regardless of infection status, providing an intriguing potential connection between expression

439   profile differences in different populations with genetic divergence between lakes. This result is

19

440 consistent with previous evidence for population structure (*e.g.* Neiman *et al.* 2004; Neiman *et al.*

441 2011, Paczesniak *et al.* 2013) and/or negative frequency-dependent selection driving divergence

442 among populations in the *P. antipodarum* system (*e.g.*, Dybdahl & Lively 1998). Together, these

443 results demonstrate that lake of origin has a major influence on gene expression and relative

444 genetic differentiation in *P. antipodarum* and provide the first genomic evidence consistent with

445 lake-specific interactions between *Microphallus* and *P. antipodarum*.

446

447 **Candidate loci for future research**. Our results provide a strong set of candidate loci for the

448 genomic basis of *P. antipodarum* and *Microphallus* interactions. Three of the 16 upregulated

449 immune-relevant genes are homologous to genes involved in response to trematode infection in

450 the *Biomphalaria glabrata-Schistosoma mansoni* host-parasite system (*e.g.,* Mitta *et al.* 2005;

451 reviewed in Coustau *et al.* 2015), suggesting these genes might play an important role in *P.*

452 *antipodarum* response to *Microphallus* infection. We also identified seven downregulated genes

453 with immune-related functions, including myosin light-chain kinase and fibrinogen-related

454 proteins, which have also been shown to contribute to resistance to infection in the *Biomphalaria-*

455 *Schistosoma* system (Mitta *et al.* 2005; Tennessen *et al.* 2015). Other significantly upregulated

456 genes with immune-related functions in infected snails represent further candidates for response

457 to *Microphallus* infection, while significantly downregulated genes with immune-related

458 functions have the potential to be involved in resisting infection.

459 Among the genes upregulated in infected snails, we observed functional enrichment for

460 actin, myosin, and genes with other cytoskeletal functions. Genes involved in cytoskeletal

461 function, including actin and myosin, are upregulated upon parasite exposure in the

462 *Biomphalaria-Schistosoma* system (reviewed in Coustau *et al.* 2015), indicating that these genes

20

463   may also contribute to response to parasite infection in *P. antipodarum*. Oxidative processes have

464   also been implicated in schistosome defence response in *Biomphalaria* (Adema *et al.* 2010,

465   reviewed in Coustau *et al.* 2015); a preliminary line of evidence that oxidative processes may also

466   be involved in the *P. antipodarum* trematode response is provided by significant functional

467   enrichment of genes involved in oxidative processes in infected snails (*e.g.*, NADH oxidation,

468   fatty acid beta-oxidation, energy derivation by oxidation of organic compounds).

469   The inclusive analysis revealed 21 upregulated and 16 downregulated genes with potential

470   roles in important behavioural traits (*e.g.*, foraging, locomotion, mating) (Fig. S2). These results

471   are consistent with evidence that exposure to (Soper *et al.* 2014) and infection by *Microphallus*

472   (Levri & Lively 1996; Levri 1999) affects *P. antipodarum* behaviour. In particular, infected

473   snails forage at a higher frequency than uninfected snails during the time of day when the

474   waterfowl that are *Microphallus*'s final host are active, rendering infected snails more vulnerable

475   to predation (Levri & Lively 1996). The implications are that these genes are a set of candidates

476   for potential genetic mechanisms and pathways involved in *Microphallus*-induced alterations to

477   *P. antipodarum* behaviour that could influence transmission probability. Future study of snails

478   from natural populations featuring little to no *Microphallus* infection (Lively 1987; Lively &

479   Jokela 2002) as well as manipulative experiments that allow comparisons between exposed *vs.*

480   unexposed and infected *vs.* uninfected individuals will provide valuable additional steps forward

481   by enabling differentiation between exposed but uninfected *vs.* naïve individuals.

482   Genes involved in the regulation of gene expression and ribosome structure and function

483   were overrepresented among significantly downregulated genes. These results are consistent with

484   our overall observation that infected snails had significantly more downregulated *vs.* upregulated

485   genes, suggesting that *Microphallus* infection leads to decreased overall gene expression in *P.*

21

486   *antipodarum*. These results are strikingly similar to a recent laboratory study showing that

487   bumblebees (*Bombus terrestris*) exposed to particularly infective genotypes of a trypanosome

488   parasite (*Crithidia bombi*) downregulated more genes than unexposed bumblebees (Barribeau *et*

489   *al.* 2014). Similar results have been reported in other lab-based studies (*e.g.,* Roger *et al.* 2008;

490   Arican-Goktas *et al.* 2014). Our results thus extend to natural coevolving populations the growing

491   body of evidence that infected hosts experience systematic downregulation of gene expression.

492

493   **Summary and conclusions.** We present novel evidence for systematic but population-specific

494   genetic and gene expression responses to parasite infection in the *P. antipodarum-Microphallus*

495   coevolutionary interaction. These results are the first genome-level evidence for the type of

496   population-specific response expected under local adaptation and coevolution in this important

497   host-parasite system. We also identified genes with functions related to immune and defence

498   response and behaviour that are likely involved in the *P. antipodarum* response to *Microphallus*

499   infection, providing a set of candidate genes for involvement in the genomic basis of coevolution.

500   The targeted exploration of these genes made possible by these results will help illuminate the

501   genetic and genomic mechanisms that determine the outcome of interactions between coevolving

502   hosts and parasites in nature.

503

509   Cytometry Facility, which is a Carver College of Medicine/Holden Comprehensive Cancer

510   Center core research facility at the University of Iowa. The Facility is funded through user fees

511   and the generous financial support of the Carver College of Medicine, Holden Comprehensive

512   Cancer Center, and Iowa City Veteran's Administration Medical Center. This project was funded

513   by NSF-MCB 1122176.

514

References

516 Abbate J, Kada S, Lion S (2015) Beyond mortality: sterility as a neglected component of parasite
517   virulence. *PLoS Pathogens*, **11**, 1-10. (doi:10.1371/journal.ppat.1005229)
518 Adema CM, Hanington PC, Lun CM, Rosenberg GH, Aragon AD, Stout BA, Lennard Richard
519   ML, Gross PS, Loker ES (2010) Differential transcriptomic responses of *Biomphalaria*
520   *glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni*
521   and *Echinostoma paraensei* (Digenea, Platyhelminthes). *Molecular Immunology*, **47,** 849-
522   860. (doi:10.1016/j.molimm.2009.10.019)
523 Adema CM, Loker ES (2015) Digenean-gastropod host associations inform on aspects of specific
524   immunity in snails. *Developmental and Comparative Immunology*, **48**, 275-283.
525   (doi:10.1016/j.dci.2014.06.014)
526 Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Available:
527   http://www.bioinformatics.babraham.ac.uk?/projects/fastqc/.
528 Arican-Goktas HD, Ittiprasert W, Bridger JM, Knight M (2014) Differential spatial repositioning
529   of activated genes in *Biomphalaria glabrata* snails infected with *Schistosoma mansoni*. *PLoS*
530   *Neglected Tropical Diseases*, **8,** 1-10. (doi:10.1371/journal.pntd.0003013)
531 Barribeau SM, Sadd BM, du Plessis L, Schmid-Hempel P (2014) Gene expression differences
532   underlying genotype-by-genotype specificity in a host-parasite system. *Proceeding of the*
533   *National Academy of Sciences USA,* **111**, 3496-3501. (doi:10.1073/pnas.1318628111)
534 Bérénos C, Wegner KM, Schmid-Hempel P (2011) Antagonistic coevolution with parasites
535   maintains host genetic diversity: an experimental test. *Proceedings of the Royal Society of*
536   *London B*, **27**, 218-224. (doi:10.1098/rspb.2010.1211)
537 Bergstrom, C. T., and L. A. Dugatkin. 2016. Evolution, 2nd edition. New York: W. W. Norton and
538   Company.
539 Brunner FS, Schmid-Hempel P, Barribeau SM (2013) Immune gene expression in *Bombus*
540   *terrestris*: signatures of infection despite strong variation among populations, colonies, and
541   sister workers. *PLoS ONE* **8**, e68181. (doi:10.1371/journal.pone.0068181)
542 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009)
543   BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421. (doi:10.1186/1471-
544   2105-10-421)
545 Chomczynski P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA, and
546   proteins from cell and tissue samples. *Biotechniques*, **15**, 532-534, 536-537.

547 Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal
548   tool for annotation, visualization and analysis in functional genomics research.
549   *Bioinformatics*, **21**, 3674-3673. (doi:10.1093/bioinformatics/bti610)
550 Coustau C, Gourbal B, Duval D, Yoshino TP, Adema CM, Mitta G (2015) Advances in gastropod
551   immunity from the study of the interaction between the snail *Biomphalaria glabrata* and its
552   parasites: A review of research progress over the last decade. *Fish and Shellfish Immunology*,
553   **4**, 5-16. (doi:10.1016/j.fsi.2015.01.036)
554 de Bekker C, Ohm RA, Loreto RG, Sebastian A, Albert I, Merrow M, Brachmann A, Hughes DP
555   (2015) Gene expression during zombie ant biting behavior reflects the complexity underlying
556   fungal parasitic behavioral manipulation. *BMC Genomics*, **16**, 620. (doi:10.1186/s12864-015-
557   1812-x)
558 Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D, De Meester L
559   (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, **450**, 870-
560   873. (doi:10.1038/nature06291)
561 Dybdahl MF, Krist AC (2004) Genotypic *vs.* condition effects on parasite-driven rare advantage.
562   *Journal of Evolutionary Biology*, **17**, 967-973. (doi:10.1111/j.1420-9101.2004.00759.x)
563 Dybdahl MF, Lively CM (1996) The geography of coevolution: comparative population structures
564   for a snail and its trematode parasite. *Evolution*, **50**(6), 2264-2275.
565 Dybdahl MF, Lively CM (1998) Host-parasite coevolution: evidence for rare advantage and time-
566   lagged selection in a natural population. *Evolution*, **52**, 1057-1066. (doi:10.2307/2411236)
567 Ederli L, Dawe A, Pasqualini S, Quaglia M, Xiong L, Gehring C (2015) Arabidopsis flower
568   specific defense gene expression patterns affect resistance to pathogens. *Frontiers in Plant*
569   *Science*, **6**, 79. (doi:10.3389/fpls.2015.00079)
570 Ehrlich PR, Raven PH (1964) Butterflies and plants: a study of coevolution. *Evolution*, **18**, 586-
571   608. (doi:10.2307/2406212)
572 Gabherr MG *et al*. (2011) Full-length transcriptome assembly from RNA-Seq data without a
573   reference genome. *Nature Biotechnology*, **29**, 644-652. (doi:10.1038/nbt.1883)
574 Gleichsner AM, Thiele EA, Minchella DJ (2015) It's All about Those Bases: The Need for
575   Incorporating Parasite Genetic Heterogeneity into the Development of Schistosome Vaccines.
576   *PLoS Neglected Tropical Diseases*, **9**(6), e0003805. doi:10.1371/journal.pntd.0003805
577 Goff L, Trapnell C, Kelley D (2013) cummeRbund: Analysis, exploration, manipulation, and
578   visualization of Cufflinks high-throughput sequencing data. R Package v. 2.12.0.
579 Gordon A, Hannon G (2010) Fastx-toolkit. FASTQ/A short-read pre-processing tools
580   (unpublished) http://hannonlabcshledu/fastx_toolkit/.
581 Haas BJ *et al*. (2013) *De novo* transcript sequence reconstruction from RNA-seq, using the Trinity
582   platform for reference generation and analysis. *Nature Protocols*, **8**, 1494-1512.
583   (doi:10.1038/nprot.2013.084)
584 Hamilton WD (1980). Sex *versus* non-sex *versus* parasite. *Oikos*, **35**, 282–290. (doi
585   10.2307/3544435)
586 Herron JC, Freeman S (2014) Evolutionary Analysis, 5th edition. Pearson Education, Inc.
587   Glenview, IL.
588 Holomuzki JR, Biggs BJF (2006) Habitat-specific variation and performance trade-offs in shell
589   armature of New Zealand mudsnails. *Ecology*, **87**, 1038-1047. (doi:10.1890/0012-
590   9658(2006)87[1038:HVAPTI]2.0.CO;2)

591   Hotez PJ (2013) Pediatric tropical diseases and the world's children living in extreme poverty.
592   *Journal of Applied Research on Children*, **4**, 10.
593   Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and
594   comparing biological sequences. *Bioinformatics*, **26**, 680-682.
595   (doi:10.1093/bioinformatics/btq003)
596   Jokela J, Dybdahl MF, Lively CM (2009) The maintenance of sex, clonal dynamics, and host-
597   parasite coevolution in a mixed population of sexual and asexual snails. *American Naturalist*,
598   **174**, 43-S53. (doi:10.1086/599080)
599   Jurberg AD, Brindley PJ (2015) Gene function in schistosomes: recent advances toward a cure.
600   *Frontiers in Genetics*, **6**, 144.(doi:10.3389/fgene.2015.00144)
601   King KC, Delph LF, Jokela J, Lively CM (2009) The geographic mosaic of sex and the Red
602   Queen. *Current Biology*, **19**, 1438-1441. (doi:10.1016/j.cub.2009.06.062)
603   King KC, Jokela J, Lively CM (2011) Parasites, sex, and clonal diversity in natural snail
604   populations. *Evolution*, **65**, 1474-1481. (doi:10.1111/j.1558-5646.2010.01215.x)
605   King KC, Jokela J, Lively CM (2011) Trematode parasites infect or die in snail hosts. *Biology
606   Letters*, **7**, 265-268. (doi:10.1098/rsbl.2010.0857)
607   King KC, Lively CM (2009) Geographic variation in sterilizing parasite species and the Red
608   Queen. *Oikos, 118*, 1416-1420. (doi:10.1111/j.1600-0706.2009.17476.x)
609   Kocot KM, Jeffery NW, Mulligan K, Halanych KM, Gregory TR (2015) Genome size estimates
610   for Aplacophora, Polyplacophora and Scaphopoda: small solenogasters and sizeable
611   scaphopods. *Journal of Molluscan Studies*. (doi:10.1093/mollus/eyv054)
612   Kofler R, Pandey RV, Schlötterer C (2011) Popoolation2: identifying differentiation between
613   populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435-
614   3436. (doi:10.1093/bioinformatics/btr589)
615   Koskella B, Lively CM (2009) Evidence for negative frequency-dependent selection during
616   experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution*, **63**,
617   2213-2221. (doi:10.1111/j.1558-5646.2009.00711.x)
618   Kover PX, Caicedo AL (2001) The genetic architecture of disease resistance in plants and the
619   maintenance of recombination by parasites. *Molecular Ecology*, **10**(1), 1-16. (doi:
620   10.1046/j.1365-294X.2001.01124.x)
621   Krist AC, Kay AD, Larkin K, Neiman M (2014) Response to phosphorus limitation varies among
622   lake populations of the freshwater snail *Potamopyrgus antipodarum*. *PLoS ONE*, **9**, e85845.
623   (doi:10.1371/journal.pone.0085845)
624   Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: exploring raw
625   genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage
626   plots. *Frontiers in Genetics*, **4**, 237. (doi:10.3389/fgene.2013.00237)
627   Laine AL (2009) Role of coevolution in generating biological diversity: spatially divergent
628   selection trajectories. *Journal of Experimental Botany*, **60**, 2957-2970.
629   (doi:10.1093/jxb/erp168)
630   Levri P (1999) Parasite-induced change in host behavior of a freshwater snail: parasitic
631   manipulation or byproduct of infection? *Behavioral Ecology*, **10**, 234-241.
632   (doi:10.1093/beheco/10.3.234)
633   Levri EP, Lively CM (1996) The effect of size, reproductive condition, and parasitism on foraging
634   behavior in a freshwater snail, *Potamopyrgus antipodarum. Animal Behavior*, **51**, 891-901.
635   (doi:10.1006/anbe.1996.0093)

636   Li J, Witten DM, Johnstone IM, Tibshirani R (2012) Normalization, testing, and false discovery
637   rate estimation for RNA-sequencing data. *Biostatistics*, **13**, 523-538.
638   (doi:10.1093/biostatistics/kxr031)
639   Lively CM (1987) Evidence from a New Zealand snail for the maintenance of sex by parasitism,
640   *Nature*, **328**, 519-521. (doi:10.1038/328519a0)
641   Lively CM, Dybdahl MF, Jokela J, Osnas EE, Delph LF (2004) Host sex and local adaptation by
642   parasites in a snail-trematode interaction. *American Naturalist*, **164**, S6-S18.
643   Lively CM, Jokela J (2002) Temporal and spatial distributions of parasites and sex in a freshwater
644   snail. *Evolutionary Ecology Research*, **4**, 219-226.
645   Luijckx P, Fienberg H, Duneau D, Ebert D (2013) A matching-allele model explains host
646   resistance to parasites. *Current Biology*, **23**, 1085-1088. (doi:10.1016/j.cub.2013.04.064)
647   Masri L *et al.* (2015). Host-pathogen coevolution: the selective advantage of *Bacillus thuringiensis*
648   virulence and its cry toxin genes. *PLoS Biology*, **13**, 1-30. (doi:10.1371/journal.pbio.1002169)
649   McTaggart SJ, Cézard T, Garbutt JS, Wilson PJ, Little TL (2015) Transcriptome profiling during a
650   natural host-parasite interaction. *BMC Genomics*, **16**, 643. (doi:10.1186/s12864-015-1838-0)
651   Micallef L, Rodgers P (2014) euler*APE*: Drawing area-proportional 3-Venn diagrams using
652   ellipses. *PLoS ONE*, **9**, e101717. (doi:10.1371/journal.pone.0101717)
653   Mitta G, Galinier R, Tisseyre P, Allienne JF, Girerd-Chambaz Y, Guillou F, Bouchut A, Coustau
654   C (2005) Gene discovery and expression analysis of immune-relevant genes in *Biomphalaria*
655   *glabrata* hemocytes. *Developmental and Comparative Immunology*, **29**, 393-407.
656   (doi:10.1016/j.dci.2004.10.002)
657   Neafsey DE *et al*. (2015) Genetic diversity and protective efficacy of the RTS,S/AS01 malaria
658   vaccine. *New England Journal of Medicine*, **373**, 2025-2037. (doi:10.1056/NEJMoa1505819)
659   Neiman M, Lively CM (2004) Pleistocene glaciation is implicated in the phylogeographic
660   structure of a New Zealand freshwater snail, *Potamopyrgus antipodarum. Molecular Ecology*,
661   **13**, 3085-3098. (doi:10.1111/j.1365-294X.2004.02292.x)
662   Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G (2011) Wide variation in ploidy
663   level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual
664   lineages. *Evolution*, **65**, 3202-3216. (doi:10.1111/j.1558-5646.2011.01360.x)
665   Osnas EE, Lively CM (2005) Immune response to sympatric and allopatric parasites in a snail-
666   trematode interaction. *Frontiers in Zoology*, **2**, 1-7. (doi:10.1186/1742-9994-2-8)
667   Paczesniak D, Jokela J, Larkin K, Neiman M (2013) Discordance between nuclear and
668   mitochondrial genomes in sexual and asexual lineages of the freshwater snail *Potamopyrgus*
669   *antipodarum. Molecular Ecology*, **22**, 4695-4710. (doi:10.1111/mec.12422)
670   Paterson S *et al.* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature*, **464,**
671   275-279. (doi:10.1038/nature08798)
672   Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, Emans R, Cesari IM, Cosseau C,
673   Mitta G (2008) Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma*
674   *mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS Neglected*
675   *Tropical Diseases*, **2**, 1-20. (doi:10.1371/journal.pntd.0000330)
676   Routtu J, Ebert D (2015) Genetic architecture of resistance in *Daphnia* hosts against two species
677   of host-specific parasite. *Heredity*, **114**, 241-248. (doi:10.1038/hdy.2014.97)
678   Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, Dunn CW
679   (2011) Resolving the evolutionary relationships of molluscs with phylogenomic tools *Nature*,
680   **480**, 364-367. (doi:10.1038/nature10526)

681   Soper DM, King KC, Vergara D, Lively CM (2014) Exposure to parasites increases promiscuity
682   in a freshwater snail. *Biology Letters*, **10**, 20131091. (doi:10.1098/rsbl.2013.1091)
683   Stutz WE, Schmerer M, Coates JL, Bolnick DI (2015) Among-lake reciprocal transplants induce
684   convergent expression of immune genes in threespine stickleback. *Molecular Ecology*, **24**,
685   4629-4646. (doi:10.1111/mec.13295)
686   Tennessen JA, Théron A, Marine M, Yeh JY, Rognon A, Blouin MS (2015) Hyperdiverse gene
687   cluster in snail host conveys resistance to human schistosome parasites. *PLoS Genetics*, **11**, 1-
688   21. (doi:10.1371/journal.pgen.1005067)
689   Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential
690   analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*, **3***1*,
691   46-53. (doi:10.1038/nbt.2450)
692   Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL,
693   Patcher L (2012) Differential gene and transcript expression analysis of RNA-Seq
694   experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562-578.
695   (doi:10.1038/nprot.2012.016)
696   van Houte, S., A. K. E. Ekroth, J. M. Broniewski, H. Chabas, B. Ashby, J. Bondy-Denomy, S.
697   Gandon, M. Boots, S. Paterson, A. Buckling, and E. R. Westra (2016). The diversity-
698   generating benefits of a prokaryotic adaptive immunity. *Nature*, in press,
699   doi:10.1038/nature17436
700   Vergara D, Lively CM, King KC, Jokela J (2013) The geographic mosaic of sex and infection in
701   lake populations of a New Zealand snail at multiple spatial scales. *American Naturalist*, **182**,
702   484-493. (doi:10.1086/671996)
703   Winterbourn ML (1973) Larval trematode parasitising the New Zealand species of *Potamopyrgus*
704   (Gastropoda: Hydrobiidae). *Mauri Ora*, **2**, 17-30.
705   Zachar N, Neiman M (2013) Profound effects of population density on fitness-related traits in an
706   invasive freshwater snail. *PLoS ONE*, **8**, e80067. (doi:10.1371/journal.pone.0080067)
707   Zimmer C, Emlen DJ (2013) Evolution: Making Sense of Life. Roberts and Company Publishers,
708   Inc. Greenwood Village, CO.
709

710   **Data Accessibility**

711   Raw RNA sequencing reads will be deposited on the NCBI Sequence Read Archive. The

712   annotated reference transcriptome assembly will be deposited on GenBank. Locations of sample

713   sites, parameters for transcriptome assembly and analyses, and GO annotation and enrichment

714   analyses are in the online supplementary information. The custom python script used in our

715   transcriptome pipeline is available at: github.com/jsharbrough/grabContigs.

716

717

718 **Author Contributions**

719 LB, MN designed research; LB performed research; LB, PF, KEM analysed data; LB, MN wrote

720 the manuscript; JLB, JML, MN funded the study; LB, MN, PF, KEM, JLB, JML edited

721 manuscript.

722

723 **Table and Figure Legends**
724

Table 1. Summary of gene expression and F_{ST} outlier analyses

Population	Expression analyses						F_{ST} outlier analyses			
	Upregulated		Downregulated	Total	FET p-value		Number of F_{ST} outlier SNPs	Number of transcripts	Upregulated	Downregulated
Alexandrina	267 (0.27)	<	728 (0.73)	995	$p < 0.0001$		42	38	2	5
Kaniere	53 (0.34)	<	103 (0.66)	156	$p = 0.0011$		50	43	0	0
Selfe	267 (0.52)	$\cong$	246 (0.48)	513	$p = 0.4821$		47	38	2	0
Inclusive	351 (0.25)	<	1057 (0.75)	1408	$p < 0.0001$		70	59	3	10

725

726 Table 1: Summary of inclusive and within-lake analyses. Left: number of significantly
727 upregulated, downregulated, and total number of significantly differentially expressed genes in
728 infected *vs.* uninfected snails (proportion of differentially expressed genes). Right: number of F_{ST}
729 outlier SNPs between infected and uninfected snails, number of transcripts containing outlier
730 SNPs, and the number of transcripts containing outlier SNPs that were significantly upregulated
731 or downregulated in infected *vs.* uninfected snails.
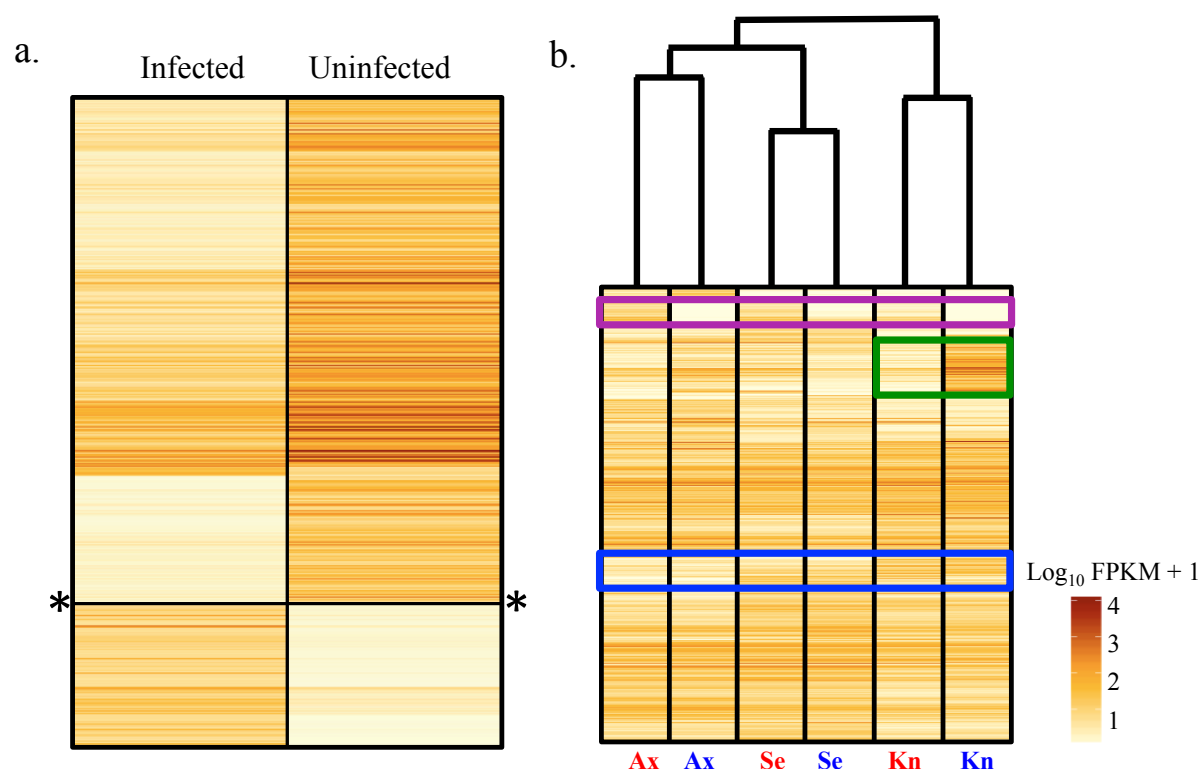732
733
734
735
736
737

28

738
739 Fig. 1. Heatmaps representing significantly differentially expressed transcripts (higher FPKM =
740 higher expression). Significance was assessed with a FDR of 5% and Benjamini-Hochberg
741 multiple test correction. a) Inclusive analysis: 1408 significantly differentially expressed
742 transcripts in infected *vs.* uninfected snails. Transcripts above the black line demarcated with "*"
743 are downregulated in infected snails (1057), and transcripts below this line are upregulated in
744 infected snails (351). b) Within-lake analysis: dendrogram and heatmap representing the 6228
745 significantly differentially expressed transcripts across infection status and lakes. "Ax" =
746 Alexandrina, "Se" = Selfe, "Kn" = Kaniere, red = infected, blue = uninfected. Snails clustered by
747 lake of origin rather than infection status, indicating spatially structured gene expression patterns.
748 Highlighted regions of heatmap contain examples of: expression differences between infected and
749 uninfected snails within a lake (green), lake-specific expression regardless of infection status
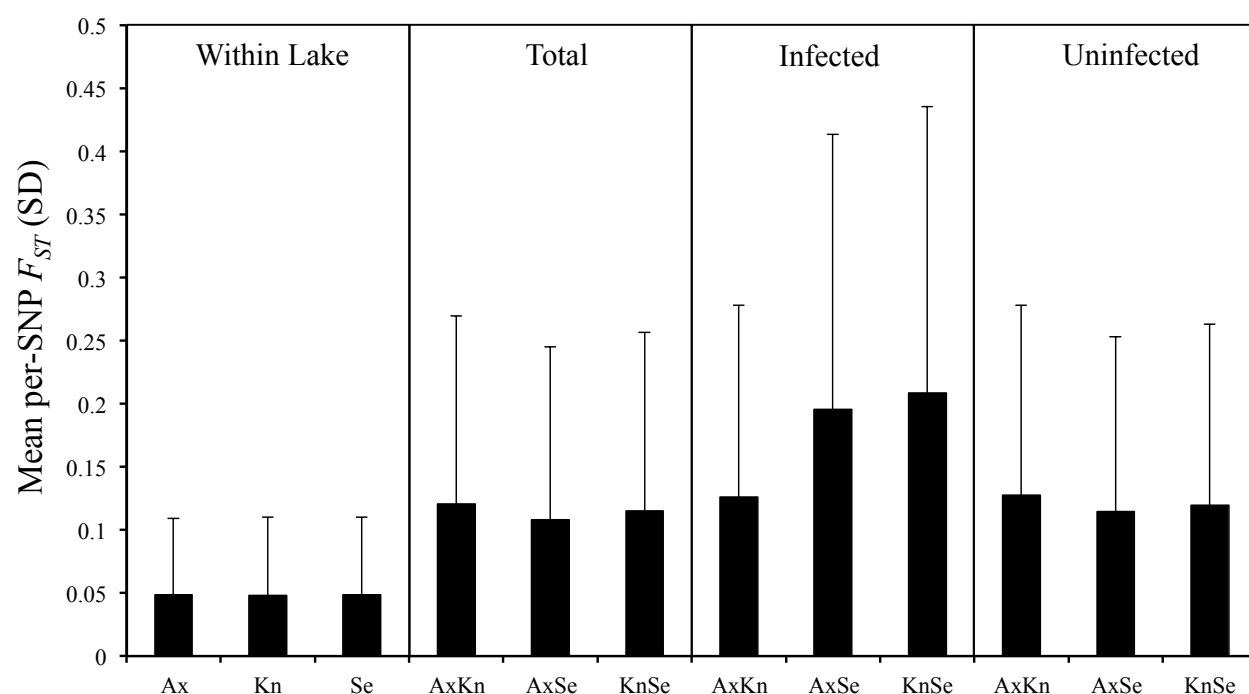750 (downregulation in Ax, blue), and across-lake upregulation in infected snails (purple).
751
752
753
754
755
756
757
758
759
760

761
762
763   Fig. 2. Mean per-SNP F_{ST} (SD) calculated based on F_{ST} per site. From left to right, the first panel
764   is mean per-SNP F_{ST} between infected and uninfected snails within each lake, the second panel is
765   mean per-SNP F_{ST} between lakes (*e.g.,* AxKn = mean per-SNP pairwise F_{ST} between Alexandrina
766   and Kaniere), the third panel is mean pairwise F_{ST} for infected replicates only, and the fourth
767   panel is mean F_{ST} for uninfected replicates only. Lake acronyms follow Fig. 1. We used Welch's
768   t-tests to compare mean F_{ST} (+/- SD) for all possible pairwise comparisons with Bonferroni
769   multiple test correction (all within-lake infected *vs.* uninfected comparisons had significantly
770   lower F_{ST} than any of the across lake comparisons; $p < 0.0001$).
771
772
773
774
775
776
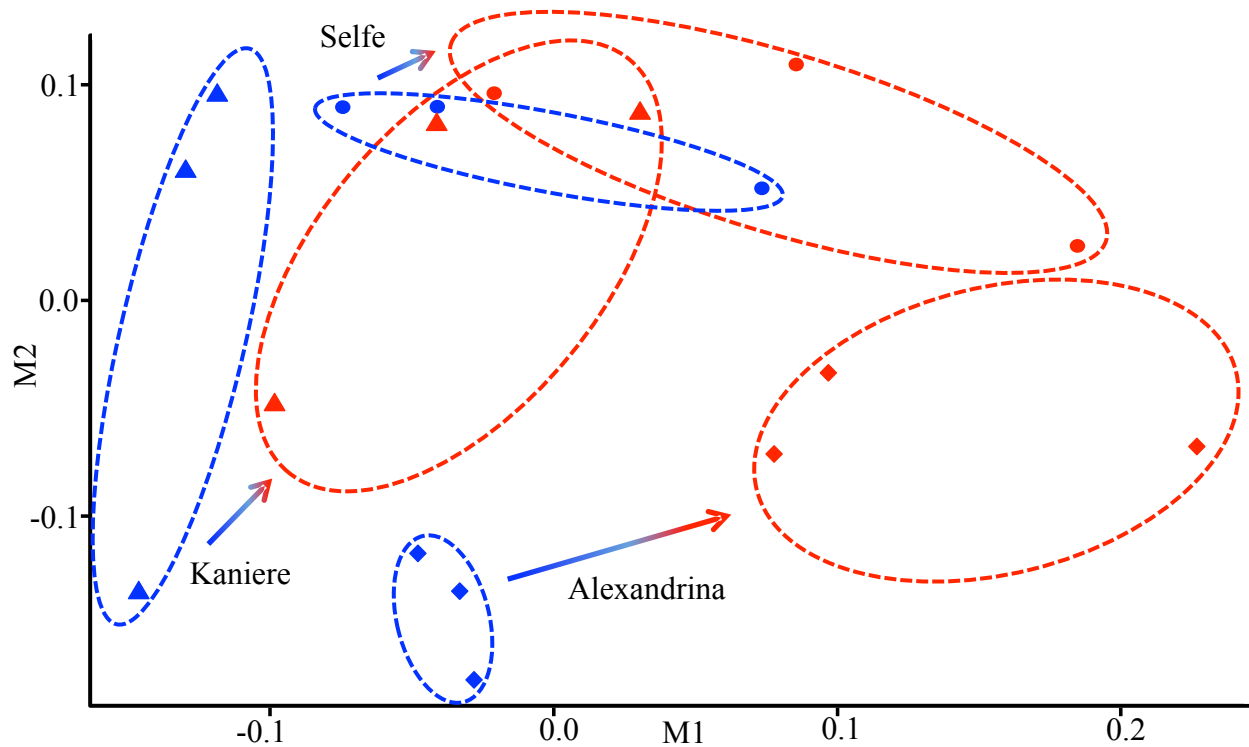777
778
779
780
781
782
783
784
785
786
787

Fig. 3. A multidimensional scaling (MDS) plot depicting how expression profiles are affected by lake population and infection status. Closer proximity of replicates in plot space indicates a more similar expression profile. Red = infected, blue = uninfected, diamonds = Alexandrina, triangles = Kaniere, circles = Selfe. The elliptical outlines (red = infected replicates from each lake; blue = uninfected replicates from each lake) delineate clustering for a particular replicate type and do not represent formal statistical tests; these clustering patterns are further supported by a dendrogram (Fig. S3). While the expression profiles that characterize uninfected and infected snails within each lake reveal lake-specific gene expression responses to infection, the consistent shift to the right along the "M1" axis for infected *vs*. uninfected snails for each lake population (highlighted with arrows) also demonstrates across-lake commonalities in expression profiles for infected snails.
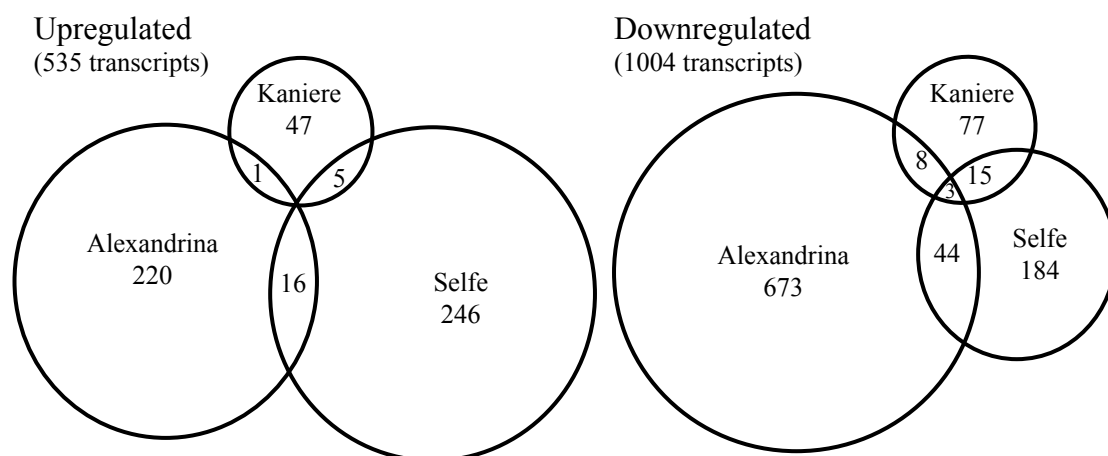
Fig. 4. Euler diagrams depicting the number of significantly differentially expressed genes in infected *vs.* uninfected snails for the within-lake analysis. Left: number of upregulated genes in single *vs.* multiple lake populations (*e.g.*, 47 genes upregulated in Kaniere; 0 genes upregulated in all three lakes). Right: number of downregulated genes in single *vs.* multiple populations. Circle size is proportional to the number of genes in that category.