Title:

Genomic evidence for population-specific responses to coevolving parasites in a New Zealand freshwater snail

Author affiliations:

Laura Bankers[1], Peter Fields[2], Kyle E. McElroy[1], Jeffrey L. Boore[3], John M. Logsdon, Jr.[1], Maurine Neiman[1]

[1]Department of Biology, University of Iowa, 143 Biology Building, Iowa City, IA 52242, USA
[2]Zoologisches Institut, Universität Basel, Vesalgasse 1, Basel, CH-4051, Switzerland
[3]Department of Integrative Biology, University of California, 3040 Valley Life Sciences Building, Berkeley, CA 94720, USA

Corresponding author*:*

Laura Bankers
laura-rice@uiowa.edu

Author Contributions*:*

LB, MN designed research; LB performed research; LB, PF, KEM analysed data; LB, MN wrote the paper; JLB, JML, MN funded the study; LB, MN, PF, KEM, JLB, JML edited manuscript.

Abstract

Reciprocal coevolving interactions between hosts and parasites are a primary source of strong selection that can promote rapid and often population- or genotype-specific evolutionary change. These host-parasite interactions are also a major source of disease. Despite their importance, very little is known about the genomic basis of coevolving host-parasite interactions, particularly in nature. Here, we use gene expression and molecular sequence evolution approaches to take critical steps towards characterizing the genomic basis of interactions between the freshwater snail *Potamopyrgus antipodarum* and its coevolving sterilizing trematode parasite, *Microphallus* sp., a textbook example of natural coevolution. We found that *Microphallus*-infected *P. antipodarum* exhibit systematic downregulation of genes relative to uninfected *P. antipodarum*. The specific genes involved in parasite response differ markedly across lakes, consistent with a scenario where population-level coevolution is leading to population-specific host-parasite interactions and evolutionary trajectories. We also identified a set of rapidly evolving loci that present promising candidates for targets of parasite-mediated selection across lakes as well as within each lake population. These results constitute the first genomic evidence for population-specific responses to coevolving infection in the *P. antipodarum-Microphallus* interaction and provide new insights into the genomic basis of coevolutionary interactions in nature.

Keywords: Host-parasite interactions, coevolution, infection, gene expression, F_{ST}

Introduction

Host-parasite interactions constitute a primary source of natural selection and provide a powerful means of evaluating the evolutionary response to strong selection (1-2). Reciprocal antagonistic

selection for greater host resistance and parasite infectivity can lead to antagonistic coevolution between host and parasite (3), which can in turn maintain high genetic diversity (4-6) and drive rapid evolutionary diversification (4, 7, 8).

Host-parasite interactions often have a genetic basis (*e.g.,* 7-9), are linked to rapid molecular evolution (*e.g.*, 7, 10), and are associated with gene expression changes (*e.g.,* 11-13). Because most studies providing important insights into the genetic and genomic basis of host-parasite coevolutionary interactions have been limited to the laboratory (*e.g.,* 11, 14-15; but also see 16-17), we know very little about the evolutionary genomics of host-parasite interactions in natural populations (17). As such, elucidating the processes of coevolution in nature is a requirement both to characterize a primary driver of selection and to formulate targeted strategies to fight natural parasite populations (*e.g.* those that infect human populations, 18). Characterization of parasite-mediated changes in host gene expression provides a powerful means of deciphering the genomic basis of natural coevolutionary interactions because expression changes are often genotype or population-specific (*e.g.,* 11), can affect host response to parasites (*e.g.,* 19), and underlie variation in resistance and susceptibility to infection (*e.g.,* 20).

The New Zealand freshwater snail *Potamopyrgus antipodarum* and its sterilizing coevolving trematode parasite *Microphallus* sp. ("*Microphallus*") provide an especially compelling context for characterizing the genomic basis of coevolution in natural populations. First, selection imposed by both host and parasite is extremely strong: infected snails are completely sterilized (21), and *Microphallus* that cannot infect the first *P. antipodarum* by which they are ingested will die (22). Second, different *P. antipodarum* populations experience consistently high *vs*. consistently low *Microphallus* infection frequencies (23-25), indicating that

47 infection frequency is a major determinant of the strength of parasite-mediated selection and,
48 thus, the rate and mode of coevolution within each population. Third, *Microphallus* is locally
49 adapted to *P. antipodarum* at both within- and among-lake scales (26-27), demonstrating the fine
50 spatial scale of coevolution in this system. Fourth, the existence of many naturally replicated and
51 separately evolving *P. antipodarum*-*Microphallus* interactions means that each population can be
52 treated as a separate evolutionary experiment into the consequences of antagonistic coevolution.
53 Fifth, rapid coevolution of *P. antipodarum* and *Microphallus* has been documented in both a
54 natural population (28) and in an experimental coevolution study (29). Finally, other snail-
55 trematode systems, many of which also involve ingestion by and infection of vertebrate final
56 hosts, are common sources of human and wildlife diseases (30-31), highlighting the translational
57 relevance of this study.

58 Multiple studies suggest that coevolutionary interactions between *P. antipodarum* and
59 *Microphallus* fit a "matching alleles" infection genetics model, whereby there are no universally
60 infective parasites or resistant hosts (26-27, 32-33). Rather, *P. antipodarum* susceptibility
61 depends on whether the genotype of the *Microphallus* individual matches the *P. antipodarum*
62 individual at loci involved in resistance. Similar matching-allele mechanisms are thought to
63 operate in other snail-trematode systems such as the laboratory model *Biomphalaria glabrata*-
64 *Schistosoma mansoni* (reviewed in 34). Recent studies suggest that both allelic identity (14) and
65 gene expression (15) are likely involved in mediating *B. glabrata* susceptibility to *S. mansoni*
66 infection, emphasizing the potentially central role for gene expression in determining outcomes
67 of host-parasite interactions in this and other snail-trematode systems (30-31).

68 Here, we take critical steps towards illuminating the genomic basis of coevolution in this
69 textbook example of antagonistic coevolution (*e.g.*, 35-36) by using RNA-Seq to perform gene

70 expression and F_{ST} analyses for three replicated samples of *Microphallus*-infected *vs.* uninfected
71 *P. antipodarum* from each of three different lake populations. In light of existing ecological
72 evidence for population structure and local adaptation in this system, we expected to 1) observe
73 divergence between populations at both the gene expression and genetic level, and 2), that such
74 divergence should be especially notable with respect to genes likely to be involved in the
75 coevolutionary response to *Microphallus* infection.
76
77 **Materials and Methods**
78 **Sample collection and dissection.** Adult *P. antipodarum* were collected from shallow water
79 (depth < 1m) habitats of three New Zealand lakes (Alexandrina, Kaniere, Selfe) known to
80 contain relatively high *Microphallus* infection frequencies (~10-20%, 25; Table S1). Following
81 transfer to the University of Iowa, we housed snails in 15 L tanks at 16°C with a 16:8 hour
82 light:dark cycle and supplied *ad libitum* dried *Spirulina* (*e.g.*, 37) until dissection.
83 Because *P. antipodarum* is polymorphic for reproductive mode and ploidy level (38, 39),
84 we confined our RNA-Seq analyses to diploid adult non-brooding (non-reproductively active)
85 females. These criteria ensured that we limited extraneous biological processes that may conflate
86 differences in gene expression related to interactions between *P. antipodarum* and *Microphallus*.
87 We dissected each snail to determine sex (male *vs.* female), *Microphallus* infection status
88 (infected *vs.* uninfected), and reproductive status (brooding *vs*. non-brooding). We used the
89 infection data to establish infection frequency for each lake sample (Table S1). Because
90 *Microphallus* infection fills the body cavity of *P. antipodarum*, we confined analyses to head
91 tissue. While the use of head tissue necessarily prevents the analysis of genes that are solely
92 expressed in body tissue, this approach is the only way to ensure that comparable tissue types are

93 isolated from both infected and uninfected snails. We stored one half of the head tissue in
94 RNA*later*® until RNA extraction and used the other head half for flow cytometric determination
95 of ploidy level (following 40; SI Methods). For each RNA-sequencing replicate, we pooled head
96 tissue from seven snails to obtain a sufficient amount of tissue for RNA extraction and
97 sequencing. We obtained three biological replicates of parasite-infected and uninfected *P.*
98 *antipodarum* from each of the three lakes, for a total of 18 replicates and 126 snails.
99
100 **RNA sequencing and *de novo* reference transcriptome assembly and annotation.** We
101 extracted RNA following the Invitrogen TRIzol protocol (41), used the Illumina Truseq LS
102 protocol for cDNA library preparation, and performed 2x100 bp paired-end RNA sequencing on
103 an Illumina HiSeq 2000 (Illumina, San Diego, CA, 2012) (SI Methods). Next, we used FASTX
104 Toolkit (42) and FastQC (43) to trim adapters, assess sequencing quality, and filter out low-
105 quality reads. We used Trinity v. 2.0.4 (44-45) to generate an initial *de novo* assembly, which we
106 filtered using TransDecoder (45), CD-HIT-EST (46), and Blobology (47) to produce a *de novo*
107 reference transcriptome containing 62862 assembled transcripts (SI Methods, Table S2).
108 Using blastx (48) and Blast2GO (49) (SI Methods), we obtained blastx annotations for
109 10171 transcripts and both blastx and GO annotations for 15797 transcripts. Nearly 75% of our
110 transcriptome did not receive GO annotations, meaning that the functions of the majority of
111 genes in our transcriptome cannot yet be determined. This result is unsurprising in light of a
112 distinct deficit of mollusc genome sequence data available to aid in annotation (50); though
113 molluscs are the most species-rich animal phylum after arthropods (51), only eight (< 1.3%) of
114 the 613 sequenced animal genomes available on NCBI as of February 2016 are from molluscs.
115

RNA-Seq gene expression analyses, functional enrichment, and F_{ST} outlier analyses. We used Tophat2 (52) to map RNA-Seq reads to our *de novo* reference transcriptome, followed by Cufflinks, Cuffmerge, and CuffDiff to estimate transcript abundance and quantify expression differences (53). We then visualized results with cummeRbund (54). We used our *de novo* transcriptome as our reference for all gene expression analyses. We used Blast2GO to compare the functions of differentially expressed genes and quantify the number of transcripts annotated with each GO term. We performed functional enrichment analyses and Fisher's Exact Tests as implemented in Blast2GO to identify significantly overrepresented functional groups among differentially expressed genes (SI Methods, Table S2).

We used F_{ST} outlier analyses to identify those genes that were evolving especially rapidly in infected *vs*. uninfected snails; these genes are candidates for *Microphallus*-mediated selection. We filtered our reference transcriptome to include only transcripts with fragments per kilobase per million reads mapped (FPKM) > 0 in all 18 replicates (30685 transcripts) to ensure each replicate was represented for all loci in the F_{ST} comparisons. We used Tophat2 (52) to map RNA-Seq reads to the filtered transcriptome and Picard Tools to prepare mapped reads for variant discovery (http://picard.sourceforge.net). Next, we used Samtools mpileup (55) to call SNPs from processed bam files, followed by Popoolation2 (56) to calculate F_{ST} per site. We used IBM SPSS Statistics v. 23 to perform outlier analyses and identify outlier SNPs between infected and uninfected snails (SI Methods). Finally, we compared levels of genetic differentiation (mean transcriptome-wide F_{ST}) between infected and uninfected snails within each lake and mean transcriptome-wide F_{ST} between lakes using Bonferroni-corrected Welch's t-tests as implemented within IBM SPSS Statistics v. 23.

We used three complementary analytical approaches to quantify gene expression and genetic differentiation (Fig. S1). First, for the "inclusive analysis," we identified differentially expressed genes likely to be broadly important for *P. antipodarum* response to *Microphallus* infection. We quantified expression for each transcript for all pooled replicates under both conditions (infected and uninfected) to determine expression differences between all infected *vs.* all uninfected snails (Fig. S1a). We then compared the annotated functions of differentially expressed genes and performed functional enrichment analyses to determine the types of genes broadly important for infection response. Next, we calculated mean transcriptome-wide F_{ST} and performed F_{ST} outlier analyses to measure genetic differentiation between infected and uninfected snails and identified functional annotations and expression patterns of each transcript containing one or more F_{ST} outlier SNPs.

Second, for the "within-lake analysis", we compared gene expression patterns between infected *vs.* uninfected snails from each lake to characterize local (lake level) gene expression responses to *Microphallus* infection. This analysis included pairwise comparisons between infected and uninfected snails from each lake population (Fig. S1b). We then compared the functions of differentially expressed genes between infected and uninfected snails from each lake. We also calculated mean transcriptome-wide F_{ST} and performed F_{ST} outlier analyses between infected and uninfected snails within each lake. Finally, we determined the functional annotation and expression pattern of each outlier-containing transcript.

Third, in the "across-lake analysis", we compared replicates by lake and infection status. Here, our goal was to use patterns of gene expression within and across lakes to evaluate evidence for local adaptation and identify genes likely to be locally important for response to *Microphallus* infection. We conducted every possible pairwise comparison of gene expression

between all replicates from all three lakes (Fig. S1c), allowing us to differentiate between genes that were differentially expressed in infected *vs*. uninfected replicates across the three lakes (*i.e*., evidence for population-specific infection responses) and genes expressed differently across lakes regardless of infection status (*i.e*., expression difference due to lake of origin rather than infection status). Annotation and functional enrichment analyses allowed us to identify putative functions of genes that were significantly differentially expressed between infected and uninfected snails in more than one lake population *vs.* within a single lake (SI Methods). We visualized these comparisons with Euler diagrams generated with eulerAPE v3 (57). Finally, we evaluated whether transcripts containing F_{ST} outliers between infected and uninfected snails contained outliers in multiple lake populations *vs.* only a single lake population.

Results

Inclusive analysis: Greater downregulation of transcripts in infected snails. We identified 1408 significantly differentially expressed transcripts (FDR: 5%, Benjamini-Hochberg) between *Microphallus*-infected and uninfected snails (Figs. 1a, S1a). A significantly higher proportion of these transcripts were downregulated (1057, ~75%) *vs*. upregulated (351, ~25%) in infected *vs.* uninfected snails (Fisher's Exact Test: $p < 0.0001$; Table 1, Fig. 1a), indicating that infected *P. antipodarum* experience systematic reduction in gene expression.

We obtained gene ontology (GO) annotations for 447 of these 1408 transcripts, 216 of which were significantly upregulated (hereafter "upregulated") and 231 significantly downregulated (hereafter "downregulated") in infected snails. 16 upregulated genes were annotated as involved in immune system processes (Fig. S2), making these loci our strongest candidates for direct involvement in response to *Microphallus* infection. The 10 genes with

putative brain/behaviour functions that were upregulated in infected snails (Fig. S2) are interesting in light of the well-characterized influence of *Microphallus* infection on *P. antipodarum* behaviour (*e.g.*, 58). Functional enrichment revealed that in infected snails, antigen processing and presentation were overrepresented among upregulated genes. Processes related to transcription and translation were overrepresented among downregulated genes, consistent with our observation that infected snails had significantly more downregulated genes (Table S3).

Our inclusive F_{ST} analysis, comparing genetic differentiation between infected and uninfected snails, revealed a mean transcriptome-wide F_{ST} (SD) of 6.34367e-5 (7.35667e-6) and identified 58 outlier SNPs (mean F_{ST} of outliers (SD) of 0.497 (0.28)) from 46 transcripts. Of these 46 transcripts, three were downregulated and two were upregulated in infected snails, meaning the majority of F_{ST} outlier-containing genes were not significantly differentially expressed (Table 1). We obtained functional annotations for 21 of these 46 transcripts. We annotated the three genes that contained F_{ST} outlier SNPs and were downregulated in infected snails with functions related to immune and stress responses (Tables S4, S5).

Within-lake analysis: Lake-specific responses to *Microphallus* infection. Similar to the inclusive analysis, we observed a significantly greater proportion of downregulated *vs.* upregulated transcripts in infected *P. antipodarum* in two of the three lakes (Table 1, Figs. 1b, S1b). A total of 1539 transcripts were significantly differentially expressed between infected and uninfected snails in at least one of the three pairwise tests (Fig. 1b).

Functional enrichment for the within-lake analysis revealed that various metabolic processes are overrepresented among upregulated genes in infected snails from Alexandrina (Table S3). We did not detect significant functional enrichment for genes upregulated in infected

207 snails from lakes Selfe or Kaniere. Upregulated genes in uninfected snails from Alexandrina and
208 Selfe are enriched for GO terms related to transcription and translation. Upregulated genes in
209 uninfected snails from Kaniere are enriched for pantothenate metabolism (Table S3).
210 The mean transcriptome-wide F_{ST} between infected and uninfected snails within each
211 lake was significantly different from each of the other lakes (Fig. 2), indicating population-
212 specific levels of genetic differentiation between infected and uninfected snails. We identified
213 45, 40, and 51 F_{ST} outlier SNPs between infected and uninfected snails (mean F_{ST} of outliers
214 (SD) of 0.51 (0.28), 0.43 (0.24), and 0.52 (0.25)) from Alexandrina, Kaniere, and Selfe,
215 respectively. Similar to the inclusive analysis, the vast majority of outlier-containing transcripts
216 (96%) were not significantly differentially expressed (Table 1).
217
218 **Across-lake analysis: Lake of origin strongly influences genetic differentiation and gene**
219 **expression**. We observed 6228 significantly differentially expressed transcripts within and
220 across lakes (Figs. 3, S1c). ~75% (4689) of these 6228 transcripts were significantly
221 differentially expressed across the three lakes regardless of infection status, indicating that lake
222 of origin has a markedly stronger influence on gene expression than infection status (Figs. 1b, 3,
223 4, S3). Similarly, when replicates are clustered based on expression profile, they group first by
224 lake of origin, followed by infection status. This result indicates both that population of origin is
225 a key determinant of gene expression in *P. antipodarum* (Figs. 3, S3) and that infection results in
226 predictable changes in expression (Fig. 3) that are detectable on this background of population
227 divergence. These findings are consistent with and extend to the gene expression level evidence
228 for population-specific phenotypes (40, 59) and marked population genetic structure (39, 60-61)

229 in *P. antipodarum*. These results also demonstrate that infection has marked consistent and
230 systematic consequences for gene expression in this species.

231 We used the total number of transcripts identified as significantly differentially expressed
232 between infected and uninfected snails (1539) in at least one of the three within-lake analyses
233 (Fig. 1b) to compare the number of significantly up or downregulated transcripts in multiple *vs*.
234 single populations, allowing us to determine whether snails from different lakes have similar
235 gene expression responses to *Microphallus* infection. This comparison revealed that nearly all of
236 the differentially expressed transcripts from the within-lake analysis (1447 transcripts, 94%) are
237 only significantly differentially expressed in a single lake. Only 6% (92) of the differentially
238 expressed transcripts showed significant differential expression in more than one population
239 (Fig. 4). In summary, the vast majority of differentially expressed transcripts between infected
240 and uninfected snails show significant up or downregulation in only one population, suggesting a
241 distinct local (lake specific) gene expression response to parasite infection.

242 Next, we identified the types of genes that were significantly differentially expressed in
243 single *vs*. multiple populations to parse out general *vs.* lake-specific infection responses. Of the
244 92 genes that were significantly differentially expressed in more than one lake, 22 genes were
245 upregulated in infected snails (Table S6). We obtained GO annotations for seven of these genes,
246 which had functions related to immune response, nervous system function, and metabolism. 10
247 of the 70 genes found to be significantly downregulated in infected snails in more than one lake
248 received GO annotations, which included immune function, response to stimulus, and
249 transcription/translation (Table S6). 396 of the 1447 genes that were significantly differentially
250 expressed in only one lake received GO annotations (Fig. S2). Even though the particular genes
251 experiencing differential expression differ on a lake-by-lake basis, these different genes often

252 belong to similar GO categories (*e.g.*, immune system processes, response to stimulus, and
253 behaviour).

254 We also evaluated how many transcripts contained F_{ST} outlier SNPs in single *vs.* multiple
255 populations. Similar to our gene expression analyses, we found that very few transcripts (6
256 transcripts, ~4%) containing F_{ST} outlier SNPs between infected and uninfected snails in one
257 population also contained outlier SNPs in another population (3 transcripts each between
258 Alexandrina and Kaniere and between Kaniere and Selfe). This result indicates that the focal loci
259 of *Microphallus*-mediated selection are likely to often be population specific. Of the transcripts
260 that contained outlier SNPs, 14, 12, and 15 transcripts were annotated for Alexandrina, Kaniere,
261 and Selfe, respectively. 10 of these transcripts were annotated as relevant to immune
262 response/function and 4 for neurological processes (Tables S4, S5).

263 Finally, we compared mean transcriptome-wide F_{ST} across lakes as a whole, for infected
264 replicates only, and for uninfected replicates only. The mean F_{ST} was significantly different ($p <$
265 0.0001, Fig. 2) in all possible pairwise comparisons. We also found that F_{ST} between infected
266 and uninfected snails within each lake was significantly lower than the mean F_{ST} for any across-
267 lake comparison (Fig. 2), indicating greater levels of genetic differentiation among lake
268 populations than between infected and uninfected snails within a population. This result is
269 consistent with and extends to the genome level the outcome of marker-based studies (*e.g.,* 61)
270 that have documented strong across-lake genetic structure in *P. antipodarum*.
271

272 **Discussion**

273 We used gene expression and F_{ST} analyses to shed light on the genomic underpinnings of
274 coevolution in a natural context. Most importantly, we found that infection of *P. antipodarum* by

its coevolving trematode parasite *Microphallus* elicits a marked, systematic, and population-specific gene expression response, with important potential implications for our understanding of the dynamics of coevolution. Our F_{ST} analyses allowed us to evaluate levels of genetic differentiation within and among lakes and revealed promising candidate genes for the focus of *Microphallus*-mediated selection. These findings provide a qualitative advance by extending evidence for local adaptation and coevolution in this textbook coevolutionary interaction to the genomic level, demonstrating distinct local genetic and gene expression responses by *P. antipodarum* to *Microphallus* infection. Together, these results illuminate the unique and important insights that can come from sampling multiple natural populations of interacting hosts and parasites.

Systematic downregulation and response to parasite infection. Our inclusive analysis revealed that the majority of transcripts that are significantly differentially expressed in infected snails are downregulated relative to uninfected snails. This pattern could reflect several non-mutually exclusive phenomena, ranging from tissue/organ destruction and/or overall poor condition of infected snails (*e.g.,* 11) to reallocation of resources to genes needed for defence against and response to *Microphallus* infection (*e.g.,* 62) to suppression of *P. antipodarum* gene expression by *Microphallus* as a means of evading host immune and defence systems (*e.g.,* 11, 13). Regardless of the specific mechanism(s) involved, the transcripts that are differentially expressed between infected and uninfected snails in the inclusive analysis represent a set of genes that are most likely to be of general importance to the *P. antipodarum*-*Microphallus* interaction. Our inclusive F_{ST} analyses yielded additional candidate genes for the focus of parasite-mediated selection, with the 58 SNPs from 46 genes that are evolving especially rapidly

298 providing a particularly strong set of candidate loci for *Microphallus* response. The presence of
299 multiple genes annotated with immune/stress responses and with neurological and behaviour-
300 related functions strengthens this conclusion.
301
302 **Genomic evidence for population-level responses to infection**. Our results demonstrate that *P.*
303 *antipodarum* response to *Microphallus* infection is largely population specific. For example, in
304 Alexandrina and Kaniere, the majority of significantly differentially expressed transcripts are
305 downregulated in infected relative to uninfected snails (73% for Alexandrina and 66% for
306 Kaniere). By contrast, the proportion of significantly upregulated and downregulated transcripts
307 is much more similar in snails from Selfe (52% *vs*. 58%, respectively). This pattern is interesting
308 in light of data from Vergara *et al.* (2013) suggesting that lake Selfe *P. antipodarum* have
309 historically experienced lower frequencies of *Microphallus* infection (3.5x lower than
310 Alexandrina, 1.4x lower than Kaniere; Table S1)(25), suggesting that lake Selfe snails are likely
311 to have experienced a somewhat weaker intensity of *Microphallus*-mediated selection.
312 The vast majority (~87%) of differentially expressed transcripts were only significantly
313 differentially expressed in a single lake population, and ~75% of significantly differentially
314 expressed transcripts were expressed differently across lakes regardless of infection status. We
315 also found significantly lower levels of genetic differentiation between snails within *vs.* across
316 lakes, regardless of infection status. This result is consistent with previous evidence for
317 population structure (*e.g.* 39, 60-61) and/or negative frequency-dependent selection driving
318 divergence between populations (*e.g.*, 63). Together, these results demonstrate that lake of origin
319 has a major influence on gene expression and genetic differentiation and provide the first
320 genomic evidence consistent with local adaptation between *Microphallus* and *P. antipodarum*.

Candidate loci for future research. Our results provide a strong set of candidate loci for the genomic basis of *P. antipodarum* and *Microphallus* interactions. Three of the 16 upregulated immune-relevant genes are homologous to genes involved in response to trematode infection in the *Biomphalaria glabrata*-*Schistosoma mansoni* host-parasite system (*e.g.,* 64; reviewed in 65), suggesting these genes might play an important role in *P. antipodarum* response to *Microphallus* infection. We also identified seven downregulated genes with immune-related functions, including myosin light-chain kinase and fibrinogen-related proteins, which have also been shown to contribute to resistance to infection in the *Biomphalaria*-*Schistosoma* system (10, 64). Other significantly upregulated genes with immune-related functions in infected snails represent further candidates for response to *Microphallus* infection, while significantly downregulated genes with immune-related functions have the potential to be involved in resisting infection.

Among the genes upregulated in infected snails, we observed functional enrichment for actin, myosin, and genes with other cytoskeletal functions. Genes involved in cytoskeletal function, including actin and myosin, are upregulated upon parasite exposure in the *Biomphalaria*-*Schistosoma* system (reviewed in 65), indicating that these genes may also contribute to response to parasite infection in *P. antipodarum*. Oxidative processes have also been implicated in schistosome defence response in *Biomphalaria* (66, reviewed in 65); a preliminary line of evidence that oxidative processes may also be involved in the *P. antipodarum* trematode response is provided by significant functional enrichment of genes involved in oxidative processes (*e.g.*, NADH oxidation, fatty acid beta-oxidation, energy derivation by oxidation of organic compounds) in infected snails.

343 The inclusive analysis revealed 21 upregulated and 16 downregulated genes with
344 potential roles in important behavioural traits (*e.g.*, foraging, locomotion, mating) (Fig. S2).
345 These results are consistent with evidence that exposure to (67) and infection by *Microphallus*
346 (58, 68) affects *P. antipodarum* behaviour. In particular, infected snails forage at a higher
347 frequency than uninfected snails during the time of day when the waterfowl that are
348 *Microphallus*'s final host are active, rendering infected snails more vulnerable to predation (68).
349 The implications are that these genes are a set of candidates for potential genetic mechanisms
350 and pathways involved in *Microphallus*-induced alterations to *P. antipodarum* behaviour that
351 could influence transmission probability. Future study of snails from natural populations
352 featuring little to no *Microphallus* infection (23, 38) as well as manipulative experiments that
353 allow comparisons between exposed *vs.* unexposed and infected *vs.* uninfected individuals will
354 provide valuable additional steps forward by enabling differentiation between exposed but
355 uninfected *vs.* naïve individuals.
356 Genes involved in the regulation of gene expression and ribosome structure and function
357 were overrepresented among significantly downregulated genes. These results are consistent
358 with our overall observation that infected snails had significantly more downregulated *vs*.
359 upregulated genes, suggesting that *Microphallus* infection leads to decreased overall gene
360 expression in *P. antipodarum*. These results are strikingly similar to a recent laboratory study
361 showing that bumblebees (*Bombus terrestris*) exposed to particularly infective genotypes of a
362 trypanosome parasite (*Crithidia bombi*) downregulated more genes than unexposed bumblebees
363 (11). Similar results have been reported in other lab-based studies (*e.g.,* 14-15). Our results thus
364 extend to natural coevolving populations the growing body of evidence that infected hosts
365 experience systematic downregulation of gene expression.

Summary and conclusions. We present novel evidence for systematic but population-specific genetic and gene expression responses to parasite infection in the *P. antipodarum-Microphallus* coevolutionary interaction. These results are the first genome-level evidence for the type of population-specific response expected under local adaptation and coevolution in this important host-parasite system. We also identified genes with functions related to immune and defence response and behaviour that are likely involved in the *P. antipodarum* response to *Microphallus* infection, providing a set of candidate genes for involvement in the genomic basis of coevolution. The targeted exploration of these genes made possible by these results will help illuminate the genetic and genomic mechanisms that determine the outcome of interactions between coevolving hosts and parasites in nature.

Data accessibility: All RNA-Seq data will be made publically available on the NCBI GEO database upon article acceptance.

Competing interests: We have no competing interests.

Acknowledgements

We thank Gery Hehman for RNA sequencing assistance, Curt Lively and Daniela Vergara for field collections, Joel Sharbrough for the custom python script used in our transcriptome pipeline, and Kayla King for helpful comments on an earlier version of this manuscript. The ploidy identification data presented herein were obtained at the Flow Cytometry Facility, which is a Carver College of Medicine/Holden Comprehensive Cancer Center core research facility at

389 the University of Iowa. The Facility is funded through user fees and the generous financial

390 support of the Carver College of Medicine, Holden Comprehensive Cancer Center, and Iowa

391 City Veteran's Administration Medical Center. This project was funded by NSF-MCB 1122176.

References

1. Hamilton WD (1980). Sex *versus* non-sex *versus* parasite. *Oikos 35*:282–290. (doi 10.2307/3544435)
2. Abbate J, Kada S, Lion S (2015) Beyond mortality: sterility as a neglected component of parasite virulence. *PLoS Pathog 11*:1-10. (doi:10.1371/journal.ppat.1005229)
3. Ehrlich PR, Raven PH (1964) Butterflies and plants: a study of coevolution. *Evolution 18*:586-608. (doi:10.2307/2406212)
4. Laine AL (2009) Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *J Exp Bot 60*:2957-2970. (doi:10.1093/jxb/erp168)
5. Bérénos C, Wegner KM, Schmid-Hempel P (2011) Antagonistic coevolution with parasites maintains host genetic diversity: an experimental test. *Proc R Soc Lond B 278*:218-224. (doi:10.1098/rspb.2010.1211)
6. King KC, Jokela J, Lively CM (2011) Parasites, sex, and clonal diversity in natural snail populations. *Evolution 65*:1474-1481. (doi:10.1111/j.1558-5646.2010.01215.x)
7. Paterson S *et al.* (2010) Antagonistic coevolution accelerates molecular evolution. *Nature 464*:275-279. (doi:10.1038/nature08798)
8. Masri L *et al.* (2015). Host-pathogen coevolution: the selective advantage of *Bacillus thuringiensis* virulence and its cry toxin genes. *PLoS Biol 13*:1-30. (doi:10.1371/journal.pbio.1002169)
9. Luijckx P, Fienberg H, Duneau D, Ebert D (2013) A matching-allele model explains host resistance to parasites. *Curr Biol 23*:1085-1088. (doi:10.1016/j.cub.2013.04.064)
10. Tennessen JA, Théron A, Marine M, Yeh JY, Rognon A, Blouin MS (2015) Hyperdiverse gene cluster in snail host conveys resistance to human schistosome parasites. *PLoS Genet 11*:1-21. (doi:10.1371/journal.pgen.1005067)
11. Barribeau SM, Sadd BM, du Plessis L, Schmid-Hempel P (2014) Gene expression differences underlying genotype-by-genotype specificity in a host-parasite system. *Proc Natl Acad Sci USA 111*:3496-3501. (doi:10.1073/pnas.1318628111)
12. McTaggart SJ, Cézard T, Garbutt JS, Wilson PJ, Little TL (2015) Transcriptome profiling during a natural host-parasite interaction. *BMC Genomics 16*:643. (doi:10.1186/s12864-015-1838-0)
13. de Bekker C, Ohm RA, Loreto RG, Sebastian A, Albert I, Merrow M, Brachmann A, Hughes DP (2015) Gene expression during zombie ant biting behavior reflects the complexity underlying fungal parasitic behavioral manipulation. *BMC Genomics 16*:620. (doi:10.1186/s12864-015-1812-x)
14. Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, Emans R, Cesari IM, Cosseau C, Mitta G (2008) Controlled chaos of polymorphic mucins in a metazoan

parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). *PLoS Negl Trop Dis 2*:1-20. (doi:10.1371/journal.pntd.0000330)

15. Arican-Goktas HD, Ittiprasert W, Bridger JM, Knight M (2014) Differential spatial repositioning of activated genes in *Biomphalaria glabrata* snails infected with *Schistosoma mansoni*. *PLoS Negl Trop Dis 8*:1-10. (doi:10.1371/journal.pntd.0003013)
16. Decaestecker E, Gaba S, Raeymaekers JAM, Stoks R, Van Kerckhoven L, Ebert D, De Meester L (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature 450*:870-873. (doi:10.1038/nature06291)
17. Routtu J, Ebert D (2015) Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasite. *Heredity 114*:241-248. (doi:10.1038/hdy.2014.97)
18. Neafsey DE *et al*. (2015) Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. *N Engl J Med 373*:2025-2037. (doi:10.1056/NEJMoa1505819)
19. Brunner FS, Schmid-Hempel P, Barribeau SM (2013) Immune gene expression in *Bombus terrestris*: signatures of infection despite strong variation among populations, colonies, and sister workers. *PLoS ONE 8*:e68181. (doi:10.1371/journal.pone.0068181)
20. Stutz WE, Schmerer M, Coates JL, Bolnick DI (2015) Among-lake reciprocal transplants induce convergent expression of immune genes in threespine stickleback. *Mol Ecol 24*:4629-4646. (doi:10.1111/mec.13295)
21. Winterbourn ML (1973) Larval trematode parasitising the New Zealand species of *Potamopyrgus* (Gastropoda: Hydrobiidae). *Mauri Ora 2*:17-30.
22. King KC, Jokela J, Lively CM (2011) Trematode parasites infect or die in snail hosts. *Biol Lett 7*:265-268. (doi:10.1098/rsbl.2010.0857)
23. Lively CM, Jokela J (2002) Temporal and spatial distributions of parasites and sex in a freshwater snail. *Evol Ecol Res 4*:219-226.
24. King KC, Lively CM (2009) Geographic variation in sterilizing parasite species and the Red Queen. *Oikos 118*:1416-1420. (doi:10.1111/j.1600-0706.2009.17476.x)
25. Vergara D, Lively CM, King KC, Jokela J (2013) The geographic mosaic of sex and infection in lake populations of a New Zealand snail at multiple spatial scales. *Am Nat 182*:484-493. (doi:10.1086/671996)
26. Lively CM, Dybdahl MF, Jokela J, Osnas EE, Delph LF (2004) Host sex and local adaptation by parasites in a snail-trematode interaction. *Am Nat 164*:S6-S18.
27. King KC, Delph LF, Jokela J, Lively CM (2009) The geographic mosaic of sex and the Red Queen. *Curr Biol 19*:1438-1441. (doi:10.1016/j.cub.2009.06.062)
28. Jokela J, Dybdahl MF, Lively CM (2009) The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *Am Nat 174*:43-S53. (doi:10.1086/599080)
29. Koskella B, Lively CM (2009) Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution 63*:2213-2221. (doi:10.1111/j.1558-5646.2009.00711.x)
30. Hotez PJ (2013) Pediatric tropical diseases and the world's children living in extreme poverty. *J Appl Res Child 4*:10.
31. Jurberg AD, Brindley PJ (2015) Gene function in schistosomes: recent advances toward a cure. *Front Genet 6*:144.(doi:10.3389/fgene.2015.00144)
32. Dybdahl MF, Krist AC (2004) Genotypic *vs.* condition effects on parasite-driven rare advantage. *J Evol Biol 17*:967-973. (doi:10.1111/j.1420-9101.2004.00759.x)

33. Osnas EE, Lively CM (2005) Immune response to sympatric and allopatric parasites in a snail-trematode interaction. *Front Zool 2*:1-7. (doi:10.1186/1742-9994-2-8)
34. Adema CM, Loker ES (2015) Digenean-gastropod host associations inform on aspects of specific immunity in snails. *Dev Comp Immunol 48*:275-283. (doi:10.1016/j.dci.2014.06.014)
35. Zimmer C, Emlen DJ (2013) Evolution: Making Sense of Life. Roberts and Company Publishers, Inc. Greenwood Village, CO.
36. Herron JC, Freeman S (2014) Evolutionary Analysis, 5th edition. Pearson Education, Inc. Glenview, IL.
37. Zachar N, Neiman M (2013) Profound effects of population density on fitness-related traits in an invasive freshwater snail. *PLoS ONE 8*:e80067. (doi:10.1371/journal.pone.0080067)
38. Lively CM (1987) Evidence from a New Zealand snail for the maintenance of sex by parasitism, *Nature 328*:519-521. (doi:10.1038/328519a0)
39. Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G (2011) Wide variation in ploidy level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution 65*:3202-3216. (doi:10.1111/j.1558-5646.2011.01360.x)
40. Krist AC, Kay AD, Larkin K, Neiman M (2014) Response to phosphorus limitation varies among lake populations of the freshwater snail *Potamopyrgus antipodarum*. *PLoS ONE 9*:e85845. (doi:10.1371/journal.pone.0085845)
41. Chomczynski P (1993) A reagent for the single-step simultaneous isolation of RNA, DNA, and proteins from cell and tissue samples. *Biotechniques 15*: 532-534, 536-537.
42. Gordon A, Hannon G (2010) Fastx-toolkit. FASTQ/A short-read pre-processing tools (unpublished) http://hannonlabcshledu/fastx_toolkit/.
43. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Available: http://www.bioinformatics.babraham.ac.uk?/projects/fastqc/.
44. Gabherr MG *et al*. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol 29*:644-652. (doi:10.1038/nbt.1883)
45. Haas BJ *et al*. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc 8*:1494-1512. (doi:10.1038/nprot.2013.084)
46. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics 26*:680-682. (doi:10.1093/bioinformatics/btq003)
47. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet 4*:237. (doi:10.3389/fgene.2013.00237)
48. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics 10*:421. (doi:10.1186/1471-2105-10-421)
49. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics 21*:3674-3673. (doi:10.1093/bioinformatics/bti610)
50. Kocot KM, Jeffery NW, Mulligan K, Halanych KM, Gregory TR (2015) Genome size estimates for Aplacophora, Polyplacophora and Scaphopoda: small solenogasters and sizeable scaphopods. *J Molluscan Stud*. (doi:10.1093/mollus/eyv054)

51. Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, Dunn CW (2011) Resolving the evolutionary relationships of molluscs with phylogenomic tools *Nature 480*:364-367. (doi:10.1038/nature10526)
52. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol 31*:46-53. (doi:10.1038/nbt.2450)
53. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Patcher L (2012) Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. *Nat Protoc 7*:562-578. (doi:10.1038/nprot.2012.016)
54. Goff L, Trapnell C, Kelley D (2013) cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R Package v. 2.12.0.
55. Li J, Witten DM, Johnstone IM, Tibshirani R (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics 13*:523-538. (doi:10.1093/biostatistics/kxr031)
56. Kofler R, Pandey RV, Schlötterer C (2011) Popoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics 27*:3435-3436. (doi:10.1093/bioinformatics/btr589)
57. Micallef L, Rodgers P (2014) euler*APE*: Drawing area-proportional 3-Venn diagrams using ellipses. *PLoS ONE 9*:e101717. (doi:10.1371/journal.pone.0101717)
58. Levri P (1999) Parasite-induced change in host behavior of a freshwater snail: parasitic manipulation or byproduct of infection? *Behav Ecol 10*:234-241. (doi:10.1093/beheco/10.3.234)
59. Holomuzki JR, Biggs BJF (2006) Habitat-specific variation and performance trade-offs in shell armature of New Zealand mudsnails. *Ecology 87*:1038-1047. (doi:10.1890/0012-9658(2006)87[1038:HVAPTI]2.0.CO;2)
60. Neiman M, Lively CM (2004) Pleistocene glaciation is implicated in the phylogeographic structure of a New Zealand freshwater snail, *Potamopyrgus antipodarum. Mol Ecol 13*:3085-3098. (doi:10.1111/j.1365-294X.2004.02292.x)
61. Paczesniak D, Jokela J, Larkin K, Neiman M (2013) Discordance between nuclear and mitochondrial genomes in sexual and asexual lineages of the freshwater snail *Potamopyrgus antipodarum. Mol Ecol 22*:4695-4710. (doi:10.1111/mec.12422)
62. Ederli L, Dawe A, Pasqualini S, Quaglia M, Xiong L, Gehring C (2015) Arabidopsis flower specific defense gene expression patterns affect resistance to pathogens. *Front Plant Sci 6*:79. (doi:10.3389/fpls.2015.00079)
63. Dybdahl MF, Lively CM (1998) Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution 52*:1057-1066. (doi:10.2307/2411236)
64. Mitta G, Galinier R, Tisseyre P, Allienne JF, Girerd-Chambaz Y, Guillou F, Bouchut A, Coustau C (2005) Gene discovery and expression analysis of immune-relevant genes in *Biomphalaria glabrata* hemocytes. *Dev Comp Immunol 29*:393-407. (doi:10.1016/j.dci.2004.10.002)
65. Coustau C, Gourbal B, Duval D, Yoshino TP, Adema CM, Mitta G (2015) Advances in gastropod immunity from the study of the interaction between the snail *Biomphalaria glabrata* and its parasites: A review of research progress over the last decade. *Fish Shellfish Immunol* 4:5-16. (doi:10.1016/j.fsi.2015.01.036)

66. Adema CM, Hanington PC, Lun CM, Rosenberg GH, Aragon AD, Stout BA, Lennard Richard ML, Gross PS, Loker ES (2010) Differential transcriptomic responses of *Biomphalaria glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni* and *Echinostoma paraensei* (Digenea, Platyhelminthes). *Mol Immunol 47*:849-860. (doi:10.1016/j.molimm.2009.10.019)
67. Soper DM, King KC, Vergara D, Lively CM (2014) Exposure to parasites increases promiscuity in a freshwater snail. *Biol Lett 10*:20131091. (doi:10.1098/rsbl.2013.1091)
68. Levri EP, Lively CM (1996) The effect of size, reproductive condition, and parasitism on foraging behavior in a freshwater snail, *Potamopyrgus antipodarum. Animal Behavior 51*:891-901. (doi:10.1006/anbe.1996.0093)

Table and Figure Legends

Table 1: Summary of inclusive and within-lake analyses. Left: number of significantly upregulated, downregulated, and total number of significantly differentially expressed genes in infected *vs.* uninfected snails (proportion of differentially expressed genes). Right: number of F_{ST} outlier SNPs between infected and uninfected snails, number of transcripts containing outlier SNPs, and the number of transcripts containing outlier SNPs that were significantly upregulated or downregulated in infected *vs.* uninfected snails.

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

Fig. 2. Mean transcriptome-wide F_{ST} (SD) calculated based on F_{ST} per site. From left to right, the first panel is mean pairwise F_{ST} between lakes (*e.g.,* AxKn = mean pairwise F_{ST} between Alexandrina and Kaniere), the second panel is mean pairwise F_{ST} for infected replicates only, the third panel is mean F_{ST} for uninfected replicates only, and the fourth panel is mean F_{ST} between infected and uninfected snails within each lake. Lake acronyms follow Fig. 1. We used Welch's t-tests to compare mean F_{ST} (+/- SD) for all possible pairwise comparisons with Bonferroni multiple test correction ($p < 0.0001$; all comparisons were statistically significant).

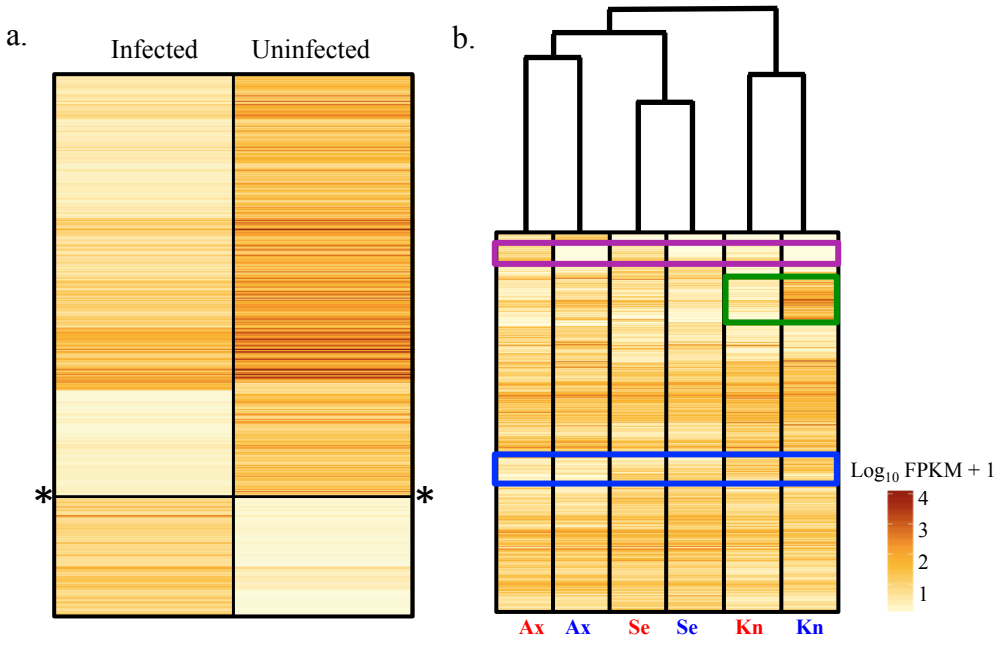
Fig. 3. A multidimensional scaling (MDS) plot depicting how expression profiles are affected by lake population and infection status. Closer proximity of replicates in plot space indicates a more similar expression profile. Red = infected, blue = uninfected, diamonds = Alexandrina, triangles = Kaniere, 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

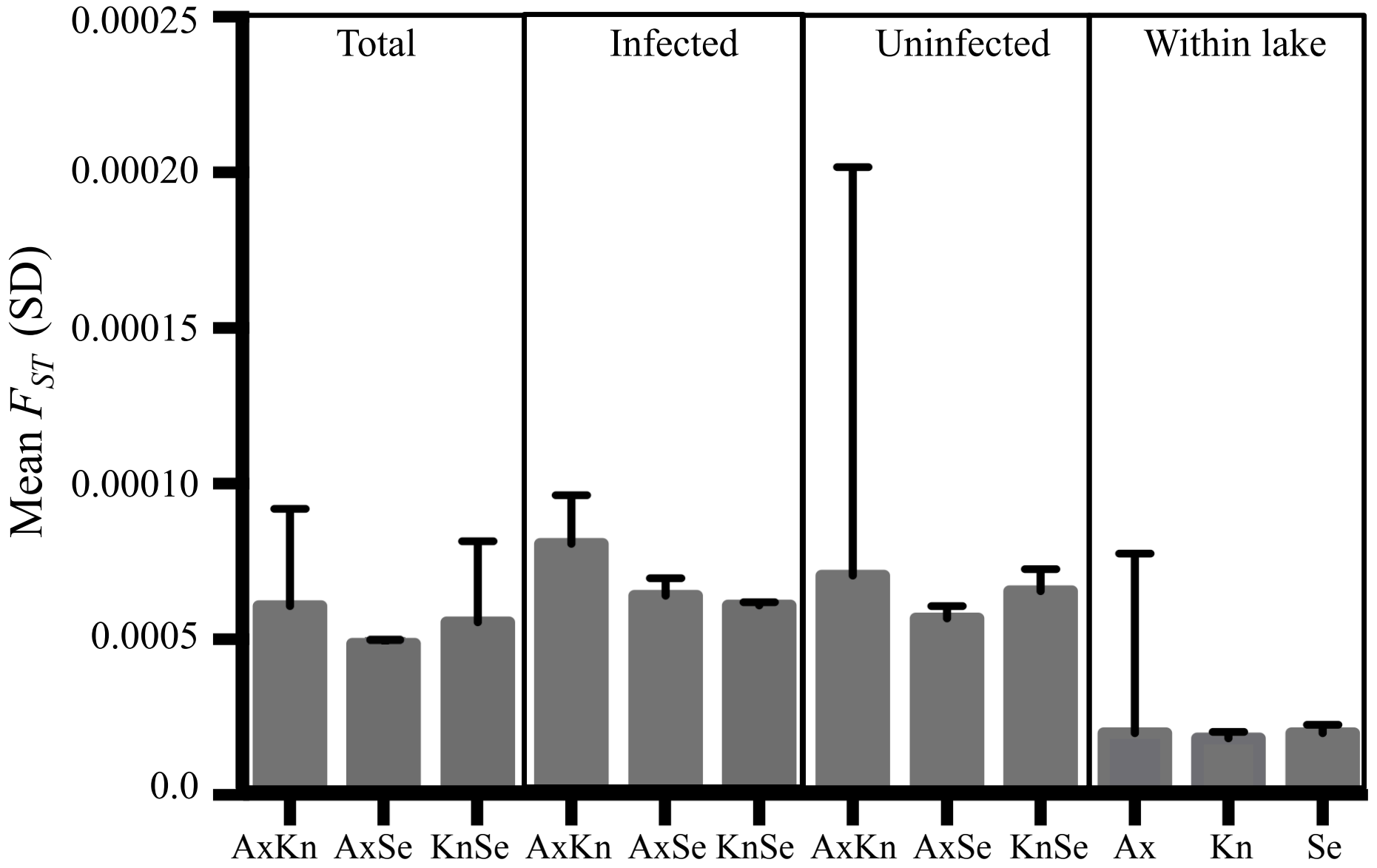
not represent formal statistical tests; these clustering patterns are further supported by a dendrogram (Fig. S3). While the expression profiles that characterize uninfected and infected snails within each lake reveal lake-specific gene expression responses to infection, the consistent shift to the right 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.

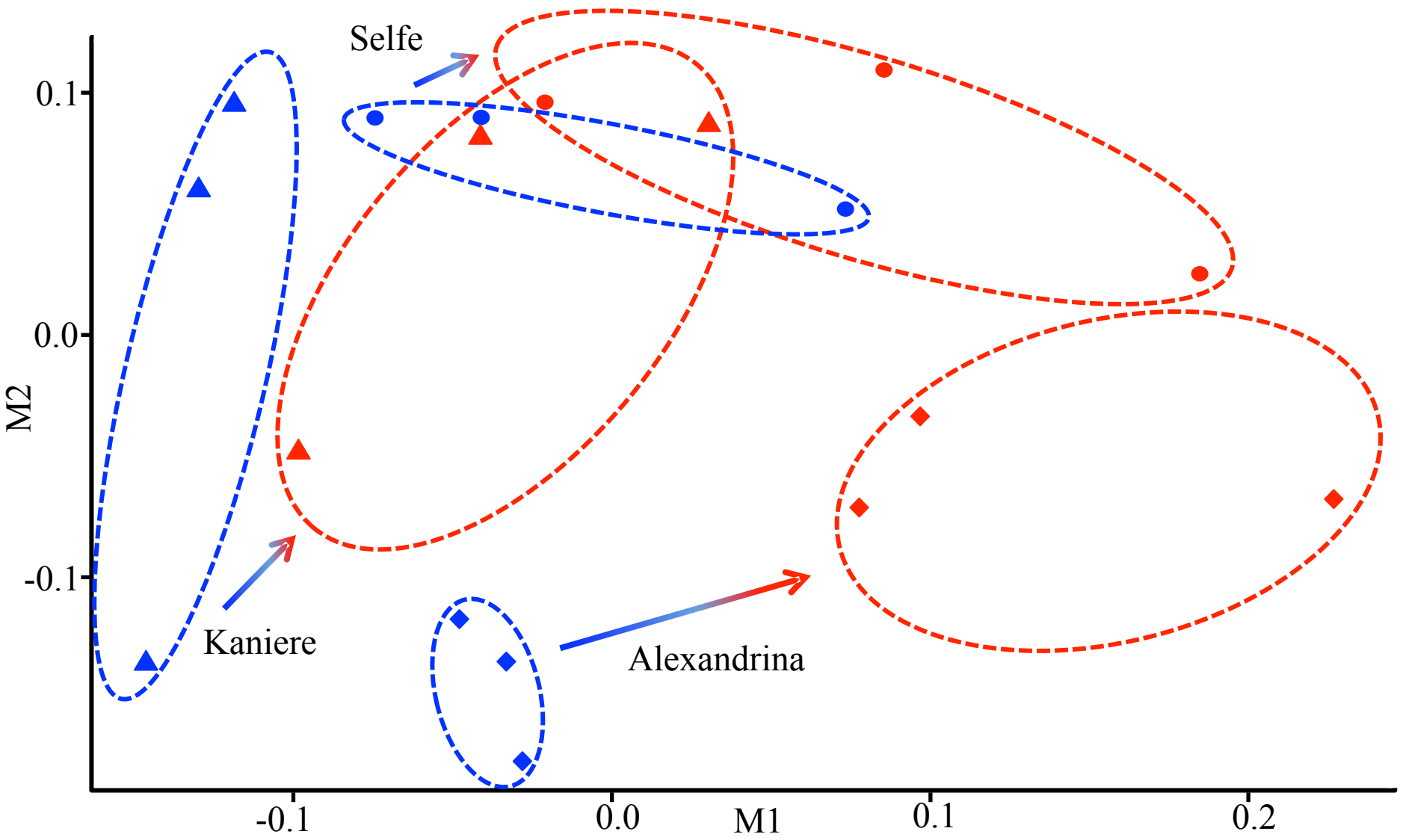
Fig. 4. Euler diagrams depicting the number of significantly differentially expressed genes in infected *vs.* uninfected snails for the within-lake analysis. Left: number of upregulated genes in single *vs*. multiple lake populations (*e.g*., 47 genes upregulated in Kaniere; 0 genes upregulated in all three lakes). Right: number of downregulated genes in single *vs*. multiple populations. Circle size is proportional to the number of genes in that category.

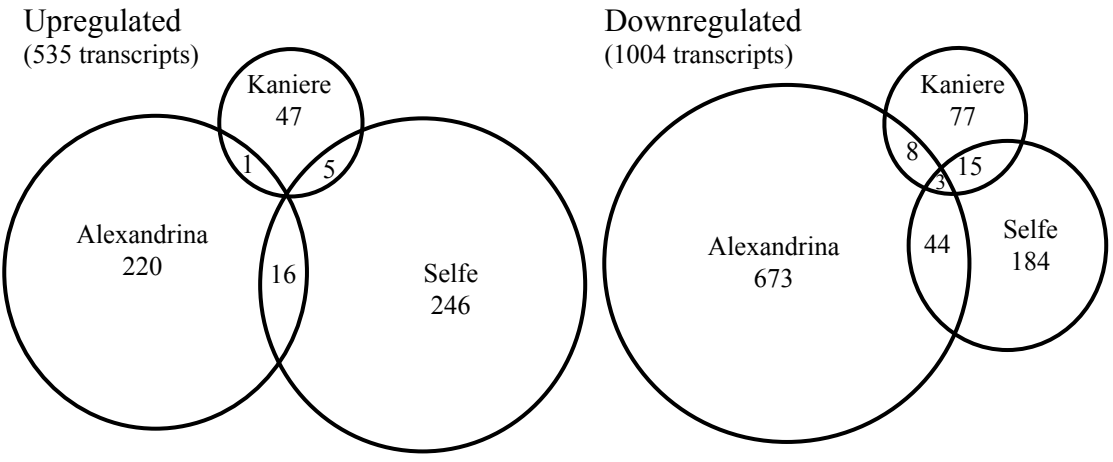
Table 1. Summary of gene expression and F_{ST} outlier analyses.

	Expression analyses					F_{ST} outlier analyses			
Population	Upregulated		Downregulated	Total	FET *p*-value	F_{ST} outliers	Transcripts	Upregulated	Downregulated
Alexandrina	267 (0.27)	<	728 (0.73)	995	$p < 0.0001$	45	37	2	6
Kaniere	53 (0.34)	<	103 (0.66)	156	$p = 0.0011$	40	31	0	0
Selfe	267 (0.52)	≅	246 (0.48)	513	$p = 0.4821$	51	35	1	1
Inclusive	351 (0.25)	<	1057 (0.75)	1408	$p < 0.0001$	58	46	1	3


a.
Infected
Uninfected
*
*
b.
Log10 FPKM + 1
4
3
2
1
Ax
Ax
Se
Se
Kn
Kn


Total
Infected
Uninfected
Within lake
Mean F_{ST} (SD)
0.00025
0.00020
0.00015
0.00010
0.0005
0.0
AxKn
AxSe
KnSe
AxKn
AxSe
KnSe
AxKn
AxSe
KnSe
Ax
Kn
Se


Selfe
Kaniere
Alexandrina
0.1
0.0
-0.1
M2
-0.1
0.0
M1
0.1
0.2


Upregulated
(535 transcripts)
Kaniere
47
1
5
Alexandrina
220
16
Selfe
246
Downregulated
(1004 transcripts)
Kaniere
77
8
15
3
Alexandrina
673
44
Selfe
184

Supplementary Methods

Sample preparation and RNA sequencing. Because *Microphallus* infection fills the body cavity of *P. antipodarum*, we confined analyses to head tissue to ensure that comparable tissue types were isolated from infected and uninfected individuals. We first separated the head tissue of dissected snails, which does not contain *Microphallus* metacercariae, from body tissue. We then split the dissected head tissue into two halves. One head half was immediately submerged in 100 µL RNA*later*® Solution (Life Technologies Corporation) and stored at 4°C for 24 hours followed by storage at -80°C (according to manufacturer protocol) until RNA extraction. The other head half was immediately snap-frozen in liquid nitrogen and stored at -80°C for flow cytometric determination of ploidy level following (40).

We extracted RNA from pooled head tissue following the Invitrogen TRIzol protocol (41). RNA quantity and quality were assessed with a Bio-Rad Experion Automated Electrophoresis Station, following manufacturer protocol for the Experion RNA analysis kit with an RQI ≥ 8 and a minimum of 2µg of total RNA. RNA shearing and cDNA library preparation were completed following the Illumina Truseq LS protocol (Illumina, San Diego, CA, 2012). Following library preparation, we used an Illumina HiSeq 2000 for 2x100 bp paired-end RNA sequencing. Each RNA sample was given a unique indexed adapter sequence (Illumina, San Diego, CA, 2012) and was then separated into two halves, allowing us to sequence each sample twice in two different lanes in order to eliminate bias due to sequencing lane. We obtained a mean read length (SD) of 100.7 bp (0.96) and a mean number of paired-end reads/replicated sample/lane (SD) of 10070420 (1474539.10), for ~20000000 paired-end reads/replicate.

***De novo* reference transcriptome assembly**. We used the FASTX Toolkit (42) and FastQC (43) to trim adapter sequences, assess sequencing quality, and remove poor-quality reads from the raw RNA-Seq data (see Table S2 for specific parameters). Quality filtering resulted in a mean number of reads/replicate/lane (SD) of 9871440.4 (1468645.04), for a total of ~18000000 paired-end reads/replicate, about twice the coverage required to quantify expression differences across a wide range of expression levels (69). We assembled a *de novo* transcriptome using Trinity v. 2.0.4 and all of the combined filtered RNA-Seq data. First, we performed the recommended Trinity *in silico* normalization for each replicate in order to reduce the amount of memory required for the assembly process while maintaining a representative read set (following 44-45). We then used Trinity to assemble the normalized reads (45), generating a transcriptome assembly with 462736 transcripts (Table S2). We annotated protein-coding regions in the transcriptome in order to identify long ORFs of putative genes that could be missed by homology searches and to filter out miscalled isoforms by using the Trinity plugin TransDecoder (45) (Table S2). This step also lessens the influence of sequencing errors, misassembled transcripts, chimeric sequences, and other common assembly issues (70). Next, we used hierarchical clustering based on sequence identity to further reduce redundancy in the transcriptome assembly, as implemented in CD-HIT-EST (46) (Table S2). Finally, we identified and eliminated potential contaminant transcripts using the 'blast_taxonomy_report.pl' script from the Blobology pipeline (47), blastx (48), and a custom python script (available upon request). We used Blobology to filter out transcripts with non-metazoan top blast hits, as well as Platyhelminthes (representing potential

47 *Microphallus* contamination). These last filtering steps provided a final reference
48 transcriptome assembly with 62862 transcripts.
49
50 **RNA-Seq gene expression analyses**. We used our *de novo* transcriptome as our
51 reference assembly for all gene expression analyses in the Tuxedo pipeline. First, we
52 mapped filtered reads to the *de novo* transcriptome assembly using TopHat2 (52). Next,
53 we assembled mapped reads into transcripts and estimated transcript abundance with
54 Cufflinks (53). We merged Cufflinks GTF and Tophat bam files for use in CuffDiff with
55 Cuffmerge (53). Finally, we identified and quantified significant changes in gene
56 expression using an FDR of 5% and Benjamini-Hochberg multiple test correction, as
57 implemented in CuffDiff (53). We required that fragments per kilobase per million reads
58 mapped (FPKM) exceeded zero to ensure that we were only comparing expression
59 patterns among genes transcribed in all replicates (Table S2). We generated heatmaps and
60 dendrograms with cummeRbund, allowing us to visualize significantly differentially
61 expressed transcripts. We also used cummeRbund to generate MDS plots (54).
62
63 **GO annotation of differentially expressed transcripts**. We annotated the reference
64 transcriptome by using blastx (48) with an E-value cutoff of 1e-5. We then imported
65 these blastx results into Blast2GO to assign gene ontology ("GO") terms to blastx-
66 annotated transcripts (49). Next, we compared the functions of differentially expressed
67 transcripts between infected and uninfected replicates overall (“inclusive analysis”) and
68 within and across lakes (“within-lake analysis” and “across-lake analysis”, respectively;
69 Fig. S1) using the combined graph function in Blast2GO. With this analysis, we

generated a list of all GO terms present in the datasets and quantified the number of transcripts annotated with each GO term. This approach allowed us to identify the putative functions of transcripts that were significantly differentially expressed between infected and uninfected replicates in more than one lake population compared to transcripts that were only significantly differentially expressed between infected and uninfected replicates in a single lake. These comparisons were visualized with Euler diagrams generated with the program eulerAPE v3 (57). Finally, we used functional enrichment analyses and Fisher's Exact Tests implemented in Blast2GO to identify significantly over *vs*. underrepresented functional groups among the differentially expressed transcripts between infected and uninfected snails in the inclusive and within-lake analyses.

SNP calling and F_{ST} outlier detection. We performed F_{ST} outlier analyses in order to identify genes evolving especially rapidly in infected *vs*. uninfected snails; these genes can be considered candidates for the focus of *Microphallus*-mediated selection. We first filtered our *de novo* reference transcriptome down to only the transcripts with an FPKM > zero in all 18 replicates (30685 transcripts) to ensure that each replicate was represented for all loci in the F_{ST} comparisons. We used Tophat2 (52) with the same parameters as in the gene expression analyses to map RNA-Seq reads to the filtered reference transcriptome. We then used Picard Tools to prepare mapped reads for variant discovery (http://picard.sourceforge.net), applying the AddOrReplaceReadGroups script to add read groups to each mapped bam file. We then merged technical replicate bam files with the MergeSamFiles script, resulting in a single bam file for each of the 18 replicates. Finally,

93 we used the MarkDuplicates script to identify duplicate reads, removing reads that
94 mapped to more than one location in the transcriptome. These processed bam files were
95 then used to generate mpileup files using Samtools mpileup (55), which calls SNPs from
96 the processed bam files (default parameters). Analyses of levels of nucleotide variation
97 from pooled RNA-sequencing data with programs like Popoolation2 (56), which was
98 developed to perform variant calling and calculate F_{ST} from pooled sequencing data, can
99 effectively identity SNPs and candidate focal genes of selection and perform F_{ST}-based
100 measures of genetic differentiation (*e.g.*, 71-72).
101 We used Popoolation2 for an inclusive analysis (all pooled infected replicates *vs.*
102 all pooled uninfected replicates), within-lake analyses (Alexandrina infected *vs.*
103 uninfected, Kaniere infected *vs.* uninfected, Selfe infected *vs.* uninfected), and across-
104 lake analyses (Alexandrina *vs.* Kaniere, Alexandrina *vs.* Selfe, Kaniere *vs.* Selfe) (Fig.
105 S1). The Popoolation2 pipeline begins by using the mpileup2sync.pl script to generate
106 synchronized mpileup files that were filtered for base quality (Q20) and then applies the
107 fst-sliding.pl script to calculate F_{ST} per site. In all cases, we required a minimum coverage
108 of 10 (to account for sequencing errors and reduce the likelihood of false positives by
109 ensuring that each SNP was sequenced multiple times), maximum coverage of 200 (to
110 reduce memory usage and biases introduced by inter-locus variation in gene expression)
111 (*e.g.*, 71-72), and otherwise used default parameters. We chose these parameters to
112 reduce the likelihood of obtaining false positives at the risk of not detecting rare alleles
113 (*e.g.,* 72). Finally, we used IBM SPSS Statistics v. 23 to perform outlier analyses to
114 identify outlier SNPs between infected and uninfected snails as a whole (inclusive
115 analysis) and between infected and uninfected snails within and between each lake and

lake pair. We also identified the expression patterns of each transcript containing F_{ST} outlier SNPs based on the expression analyses detailed above and evaluated whether transcripts containing F_{ST} outliers between infected and uninfected snails contained outliers in multiple lake populations *vs.* only a single lake population. Finally, we compared mean transcriptome-wide F_{ST} between infected and uninfected snails within each lake and mean transcriptome-wide F_{ST} between lakes using Welch's t-tests as implemented in IBM SPSS Statistics v. 23 (Bonferroni corrected).

Additional References (SI Methods only)

69. Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Mol Ecol Resour* 13:559-572. (doi:10.1111/1755-0998.12109)
70. De Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences – current advances and future possibilities. *Mol Ecol 24*:2310-2323. (doi:10.1111/mec.13165)
71. Fischer MC Rellstab C, Tedder A, Zoller S, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2013) Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Mol Ecol 22*:5594-5607. (doi:10.1111/mec.12521)
72. Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W (2014) Accuracy of allele frequency estimation using pooled RNA-Seq. *Mol Ecol Resour 14*:381-392. (doi:10.1111/1755-0998.12186)

Table S1. Location of sample sites and infection frequencies.

Lake	Location	Infection frequency (Vergara *et al.* 2013)	Infection frequency (present study)
Alexandrina	-43.936736º, 170.460260º	25.4%	20.90%
Kaniere	-42.806435º, 171.127667º	10.3% (SD +/- 6.06)	10.60%
Selfe	-43.243705º, 171.518647º	7.36% (SD +/- 6.53)	11.50%

Geographical coordinates of lakes sampled in this study and *Microphallus* infection frequencies as reported by Vergara *et al*. (2013) and in this study. Means and standard deviations are given for data from Vergara *et al*. (2013) when multiple years of sampling were available for the shallow region of a given lake (26). We determined infection frequency for this study by dissecting 116, 107, and 100 snails from Alexandrina, Kaniere, and Selfe, respectively.

Table S2. Programs and associated parameters used in transcriptome assembly and quantification and comparisons of patterns of gene expression.

Program	Purpose	Parameters	References
FASTX Toolkit	Quality filtering of raw RNA-Seq reads	Mean sequence Phred quality score cutoff = 20	53
Trinity (*in silico* normalization)	Reduces memory requirement for assembly process, recommended for large datasets	Minimum kmer coverage = 2, maximum kmer coverage = 30	55, 56
Trinity (assembly)	*De novo* assembly of RNA-Seq data using normalized read set	Default parameters	55, 56
TransDecoder	ORF annotation and miscalled isoform identification and filtering	Default parameters	56
CD-HIT-EST	Hierarchical clustering-based method to reduce assembly redundancy	Minimum sequence identity = 0.95, word size = 8 nucleotides	57
Blobology (blast taxonomy report)	Identification of potential contaminant transcripts	Default parameters	58
Tophat2	Map filtered RNA-Seq reads to *de novo* reference transcriptome	3% mismatch between reads to account for polymorphism and sequencing error	63, 64
Cufflinks	Assemble mapped reads into transcripts and estimate their abundances	Upper-quartile normalization; multi-read correction	63, 64
Cuffmerge	Merge Cufflinks gtf files and Tophat2 bam files for use in CuffDiff	Default parameters	64
CuffDiff	Quantify significant changes in gene expression	FDR = 0.05, Benjamini-Hochberg multiple test correction, FPKM > 0	64

Table S3. Functionally enriched GO terms for significantly differentially expressed genes between infected and uninfected snails for the inclusive and within lake analyses.

Functionally enriched GO terms for genes upregulated in infected snails in the inclusive analysis						
	GO-ID	Term	FDR	*p*-value	Percent in Test	Percent in Reference
	GO:0032555	purine ribonucleotide binding	7.52E-08	5.73E-11	16.52	10.68
	GO:0030554	adenyl nucleotide binding	1.56E-06	1.90E-09	13.68	8.62
	GO:0005524	ATP binding	1.80E-06	2.33E-09	13.39	8.38
	GO:0019752	carboxylic acid metabolic process	2.68E-06	3.77E-09	8.55	3.89
	GO:0015629	actin cytoskeleton	4.24E-05	8.72E-08	4.27	1.16
	GO:0005737	cytoplasm	1.17E-04	2.51E-07	28.77	29.87
	GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	1.56E-04	3.69E-07	1.71	0.10
	GO:0005829	cytosol	3.83E-04	9.92E-07	10.26	6.82
	GO:0030016	myofibril	1.61E-03	4.80E-06	2.85	0.68
	GO:0006508	proteolysis	2.53E-03	8.10E-06	6.55	3.67
	GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	3.43E-03	1.23E-05	1.42	0.11
	GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	3.43E-03	1.23E-05	1.42	0.11
	GO:0044262	cellular carbohydrate metabolic process	3.50E-03	1.39E-05	2.85	0.78
	GO:0006094	gluconeogenesis	3.50E-03	1.55E-05	1.71	0.22
	GO:0006096	glycolytic process	3.66E-03	1.64E-05	1.99	0.34
	GO:0008092	cytoskeletal protein binding	4.27E-03	2.08E-05	4.84	2.32
	GO:0022624	proteasome accessory complex	9.68E-03	5.68E-05	1.42	0.16
	GO:0004613	phosphoenolpyruvate carboxykinase (GTP) activity	1.35E-02	8.36E-05	0.85	0.03
	GO:0006521	regulation of cellular amino acid metabolic process	1.41E-02	8.79E-05	1.14	0.09
	GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.42E-02	8.99E-05	1.42	0.18
	GO:0015980	energy derivation by oxidation of organic compounds	1.75E-02	1.13E-04	2.85	1.01
	GO:0044449	contractile fiber part	1.85E-02	1.21E-04	2.28	0.64
	GO:0006116	NADH oxidation	2.63E-02	1.84E-04	0.57	0.00
	GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	3.34E-02	2.52E-04	1.42	0.23
	GO:0005977	glycogen metabolic process	3.65E-02	2.82E-04	1.42	0.24

	GO:0006635	fatty acid beta-oxidation	3.65E-02	2.86E-04	1.14	0.13
	GO:0036464	cytoplasmic ribonucleoprotein granule	4.81E-02	3.88E-04	1.42	0.26
Functionally enriched GO terms for genes upregulated in uninfected snails (downregulated in infected snails) in the inclusive analysis						
	GO-ID	Term	FDR	*p*-value	Percent in Test	Percent in Reference
	GO:0003735	structural constituent of ribosome	3.62E-10	1.40E-13	2.37	1.46
	GO:0008810	cellulase activity	9.86E-09	6.01E-12	0.66	0.01
	GO:0030247	polysaccharide binding	2.36E-08	1.98E-11	0.66	0.02
	GO:0016985	mannan endo-1,4-beta-mannosidase activity	2.10E-07	2.72E-10	0.66	0.04
	GO:0046355	mannan catabolic process	2.10E-07	2.72E-10	0.66	0.04
	GO:0006412	translation	3.07E-05	4.91E-08	2.84	3.89
	GO:0022626	cytosolic ribosome	2.80E-04	5.76E-07	1.14	0.73
	GO:0042742	defense response to bacterium	4.93E-03	1.09E-05	0.57	0.18
	GO:0030833	regulation of actin filament polymerization	1.06E-02	2.66E-05	0.66	0.32
	GO:0009506	plasmodesma	1.65E-02	4.77E-05	0.47	0.14
	GO:0045493	xylan catabolic process	2.14E-02	6.70E-05	0.28	0.02
	GO:0015935	small ribosomal subunit	2.27E-02	7.25E-05	0.76	0.53
	GO:0030301	cholesterol transport	3.34E-02	1.12E-04	0.47	0.17
	GO:0015934	large ribosomal subunit	4.86E-02	1.67E-04	0.66	0.44
Functionally enriched GO terms for genes upregulated in infected snails in the within lake analysis						
	GO-ID	Term	FDR	*p*-value	Percent in Test	Percent in Reference
Alexandrina	GO:0004613	phosphoenolpyruvate carboxykinase (GTP) activity	4.92E-03	7.50E-07	6.67	0.03
	GO:0006089	lactate metabolic process	1.76E-02	7.93E-06	4.44	0.00
	GO:0042593	glucose homeostasis	1.76E-02	8.59E-06	8.89	0.27
	GO:0010817	regulation of hormone levels	2.00E-02	1.22E-05	11.11	0.64
	GO:0030073	insulin secretion	2.22E-02	1.52E-05	8.89	0.31
	GO:0046327	glycerol biosynthetic process from pyruvate	2.84E-02	2.38E-05	4.44	0.01
Kaniere	None					
Selfe	None					
Functionally enriched GO terms for genes upregulated in uninfected snails in the within lake analysis						
	GO-ID	Term	FDR	*p*-value	Percent in Test	Percent in Reference
Alexandrina	GO:0003735	structural constituent of ribosome	2.64E-07	6.04E-11	9.63	1.51
	GO:0006412	translation	2.02E-04	7.70E-08	12.84	3.96
	GO:0022626	cytosolic ribosome	2.98E-03	1.59E-06	5.05	0.75
	GO:0016798	hydrolase activity, acting on glycosyl bonds	2.18E-02	1.65E-05	5.96	1.37
	GO:0015935	small ribosomal subunit	4.39E-02	3.68E-05	3.67	0.53

Kaniere	GO:0017159	pantetheine hydrolase activity	1.61E-02	1.23E-06	11.11	0.00
	GO:0015939	pantothenate metabolic process	2.41E-02	3.68E-06	11.11	0.01
Selfe	GO:0006418	tRNA aminoacylation for protein translation	3.26E-09	4.97E-13	13.75	0.48
	GO:0004812	aminoacyl-tRNA ligase activity	3.26E-09	4.97E-13	13.75	0.48
	GO:0000049	tRNA binding	7.01E-06	4.81E-09	7.50	0.14
	GO:0043022	ribosome binding	5.46E-04	6.23E-07	6.25	0.18
	GO:0032797	SMN complex	1.19E-02	2.53E-05	2.50	0.00
	GO:0032575	ATP-dependent 5'-3' RNA helicase activity	1.19E-02	2.53E-05	2.50	0.00
	GO:0005829	cytosol	1.40E-02	3.36E-05	21.25	6.97
	GO:0000166	nucleotide binding	2.36E-02	6.39E-05	35.00	16.75
	GO:0022403	cell cycle phase	4.06E-02	1.14E-04	3.75	0.10

Functionally enriched GO terms for the inclusive analysis and the within-lake analysis; enrichment was determined via significant outcome of Fisher's Exact tests with FDR < 0.05 as implemented in Blast2GO. "Percent in Test" refers to the percent of transcripts in the test dataset (*e.g.*, upregulated in infected snails in the inclusive analysis) that were assigned a particular GO term. "Percent in Reference" refers to the percent of transcripts in the reference transcriptome that were assigned a particular GO term.

Table S4. Functional annotation of transcripts containing at least one F_{ST} outlier SNP between infected and uninfected snails and the expression pattern of these transcripts.

Comparison	Annotation	Expression Pattern
Inclusive analysis	cathepsin l	downregulated
	probable serine threonine-protein kinase fhkb	n.s.
	cellulase egx3	n.s.
	protein deglycase dj-1	n.s.
	alkali-sensitive linkage protein 1-like	n.s.
	pancreatic triacylglycerol lipase-like	upregulated
	lambda-crystallin homolog	n.s.
	upf0764 protein c16orf89 homolog	n.s.
	PREDICTED: uncharacterized protein LOC106079933	n.s.
	PREDICTED: uncharacterized protein C9orf78 homolog	n.s.
	alpha-aminoadipic semialdehyde dehydrogenase	n.s.
	alpha-l-fucosidase-like	n.s.
	deleted in malignant brain tumors 1 partial	downregulated
	kielin chordin-like protein	n.s.
	intraflagellar transport protein 88 homolog isoform x1	n.s.
	solute carrier family 26 member 6-like	n.s.
	exonuclease 3 -5 domain-containing protein 1	n.s.
	cell wall integrity and stress response component 4-like	n.s.
	signal recognition particle 19 kda protein	n.s.
	contactin-like	n.s.
	c-type lectin	n.s.
Within lake		
Alexandrina	transcription elongation factor b polypeptide 2-like	n.s.
	chymotrypsin-like serine proteinase	downregulated
	perlucin-like protein	n.s.
	mammalian ependymin-related protein 1-like	n.s.
	uncharacterized oxidoreductase -like	n.s.
	amyloid beta a4 protein isoform x2	n.s.
	tyrosinase-like protein tyr-1	n.s.
	pleckstrin homology domain-containing family b member 2-like	n.s.
	microtubule-actin cross-linking factor isoforms 1 2 3 5-like	n.s.
	cytochrome p450 3a7-like	n.s.
	neurexin-4 isoform x2	downregulated
	complement c1q-like protein 2	n.s.
	v-type proton atpase 16 kda proteolipid subunit	n.s.
	adenylate kinase 9	n.s.

Kaniere	proteasome subunit beta type-1	n.s.
	60s ribosomal protein l23a	n.s.
	electron transfer flavoprotein subunit mitochondrial	n.s.
	chitin-binding protein	n.s.
	cd109 antigen-like	n.s.
	kielin chordin-like protein	n.s.
	hemk methyltransferase family member 1	n.s.
	ganglioside gm2 activator	n.s.
	proteasome inhibitor pi31 subunit	n.s.
	aconitate mitochondrial-like	n.s.
	ependymin-related protein 1-like	n.s.
	protein pif-like	n.s.
Selfe	39s ribosomal protein mitochondrial	n.s.
	neuropilin-2 isoform x2	n.s.
	transmembrane protease serine 2	n.s.
	goose-type lysozyme 2	n.s.
	carboxypeptidase b-like	n.s.
	thioester-containing protein	n.s.
	mucin-5ac-like isoform x1	n.s.
	peptidoglycan recognition protein s11	n.s.
	mucin-5ac- partial	n.s.
	papilin-like isoform x1	n.s.
	neurocalcin homolog	n.s.
	signal peptidase complex subunit 3-like	n.s.
	sperm-associated antigen 6	n.s.
	cartilage matrix	n.s.
	blastula protease 10	n.s.

Annotations for transcripts containing at least one F_{ST} outlier SNP between infected and uninfected snails and that received functional annotations using Blast2GO. The expression pattern of each transcript is based on the expression analyses for the analogous comparison. Upregulated means the transcript was upregulated in infected relative to uninfected snails, downregulated means that the transcript was downregulated in infected relative to uninfected snails, and "n.s." means that the transcript was not significantly differentially expressed between infected and uninfected snails.

Table S5. Level 2 GO terms assigned to annotated transcripts that also contained at least one F_{ST} outlier SNP as well as the number of transcripts that received each GO term within a particular comparison.

Comparison	GO-id	GO-term	Number of Transcripts
Inclusive analysis	GO:0051179	localization	5
	GO:0071840	cellular component organization or biogenesis	4
	GO:0023052	signaling	3
	GO:0009987	cellular process	9
	GO:0051704	multi-organism process	2
	GO:0032502	developmental process	2
	GO:0044699	single-organism process	10
	GO:0008152	metabolic process	13
	GO:0065007	biological regulation	4
	GO:0022414	reproductive process	2
	GO:0050896	response to stimulus	3
	GO:0002376	immune system process	1
	GO:0032501	multicellular organismal process	3
	GO:0007610	behavior	2
	GO:0000003	reproduction	2
Within lake			
Alexandrina	GO:0051179	localization	2
	GO:0071840	cellular component organization or biogenesis	2
	GO:0023052	signaling	1
	GO:0009987	cellular process	5
	GO:0051704	multi-organism process	2
	GO:0032502	developmental process	3
	GO:0040011	locomotion	1
	GO:0044699	single-organism process	7
	GO:0008152	metabolic process	8
	GO:0065007	biological regulation	2
	GO:0022610	biological adhesion	2
	GO:0022414	reproductive process	1
	GO:0050896	response to stimulus	2
	GO:0040007	growth	1
	GO:0007610	behavior	1
	GO:0032501	multicellular organismal process	3
	GO:0000003	reproduction	1
Kaniere	GO:0051179	localization	1
	GO:0009987	cellular process	8
	GO:0044699	single-organism process	3
	GO:0008152	metabolic process	10

	GO:0065007	biological regulation	5
	GO:0022610	biological adhesion	1
	GO:0007610	behavior	1
	GO:0032501	multicellular organismal process	1
Selfe	GO:0051179	localization	3
	GO:0071840	cellular component organization or biogenesis	3
	GO:0023052	signaling	3
	GO:0051704	multi-organism process	2
	GO:0009987	cellular process	9
	GO:0032502	developmental process	2
	GO:0040011	locomotion	2
	GO:0044699	single-organism process	5
	GO:0008152	metabolic process	11
	GO:0065007	biological regulation	7
	GO:0022610	biological adhesion	2
	GO:0050896	response to stimulus	6
	GO:0040007	growth	1
	GO:0032501	multicellular organismal process	3

Table S6. Genes with significant up *vs*. downregulation in more than one population in the within-lake analysis that received both blastx and GO annotations.

	Upregulated in infected snails			Downregulated in infected snails		
Populations	Total # of transcripts	# of transcripts annotated	Annotations	Total # of transcripts	# of transcripts annotated	Annotations
Alexandrina and Kaniere	1	0	N/A	8	2	ribosomal protein l44e uncharacterized oxidoreductase yrbe-like
Alexandrina and Selfe	15	6	protein jagged-2 tryptophan -dioxygenase annexin a7 tyramine beta-hydroxylase phosphoenolpyruvate cytosolic phosphoenolpyruvate carboxykinase	44	11	tyrosinase-like protein tyr-3 achain keyhole limpet hemocyanin hemocyanin isoform 1 beta- -galactosyl-o-glycosyl-glycoprotein beta- -n-acetylglucosaminyltransferase 4 von willebrand factor type a 5 -nucleotidase elongation factor 2 gtp-binding protein 1 lysine--trna ligase-like eukaryotic translation initiation factor 2 subunit 3-like low quality protein: hemicentin-1
Kaniere and Selfe	5	0	N/A	15	2	myosin heavy chain collagen alpha-1 chain-like
Alexandrina, Kaniere, and Selfe	0	0	N/A	3	0	N/A

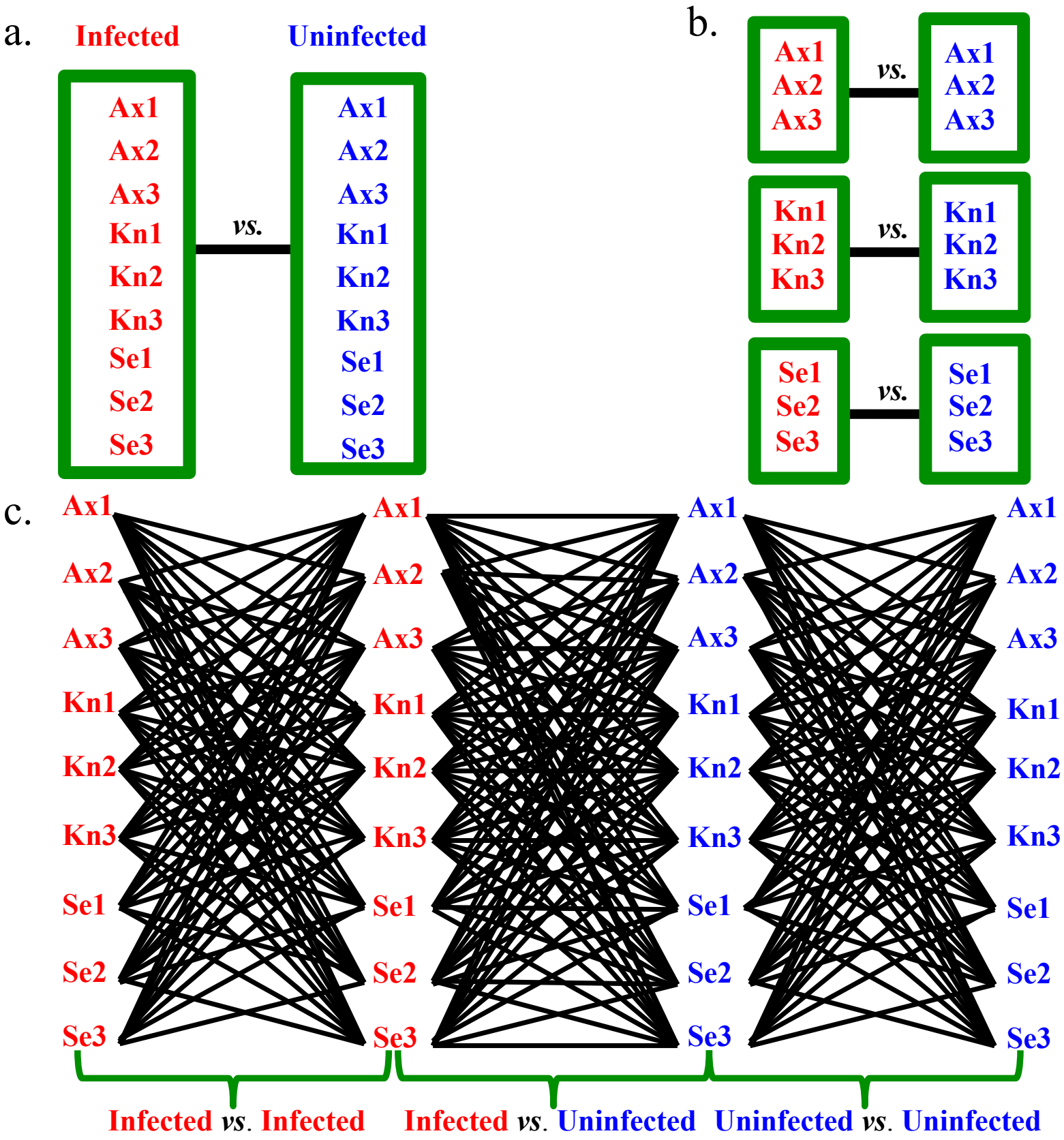


Fig. S1. Schematic depicting our three hierarchical analyses. Ax = Alexandrina, Kn = Kaniere, Se = Selfe; numbers after lake names refer to biological replicates. Black lines are drawn between the groups/samples being compared. a) "inclusive analysis," comparing all infected samples pooled *vs.* all uninfected samples pooled, b) "within-lake analysis," comparing infected *vs.* uninfected samples within each lake, c) "across-lake analysis," performing all possible pairwise comparisons between samples.

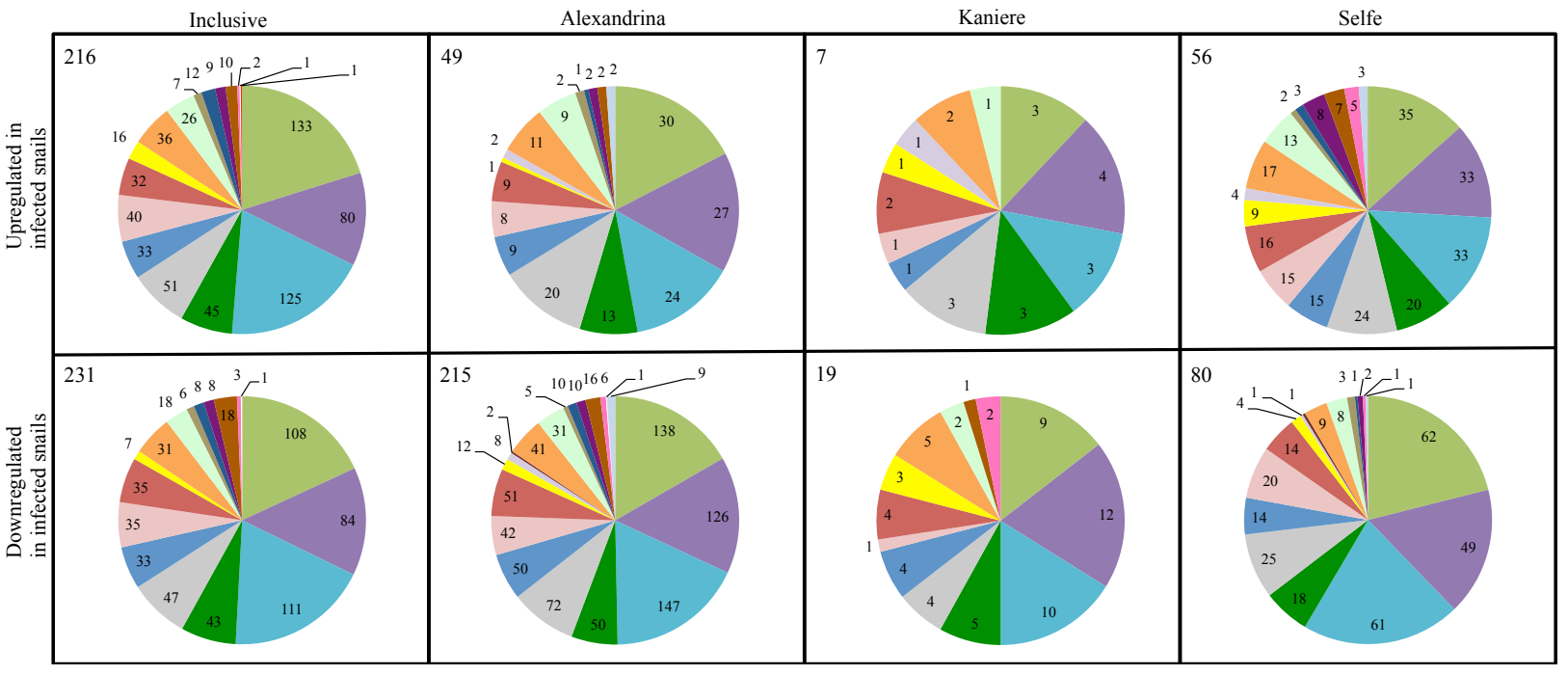


Fig. S2. Distribution of GO terms identified by Blast2GO for genes that showed significant differential expression; numbers indicate the number of genes annotated within each particular GO category. From left to right: inclusive analysis, and the three within-lake analyses. Many transcripts are represented by more than one GO term. The number of annotated transcripts that each pie chart represents is in the top left corner of each box.

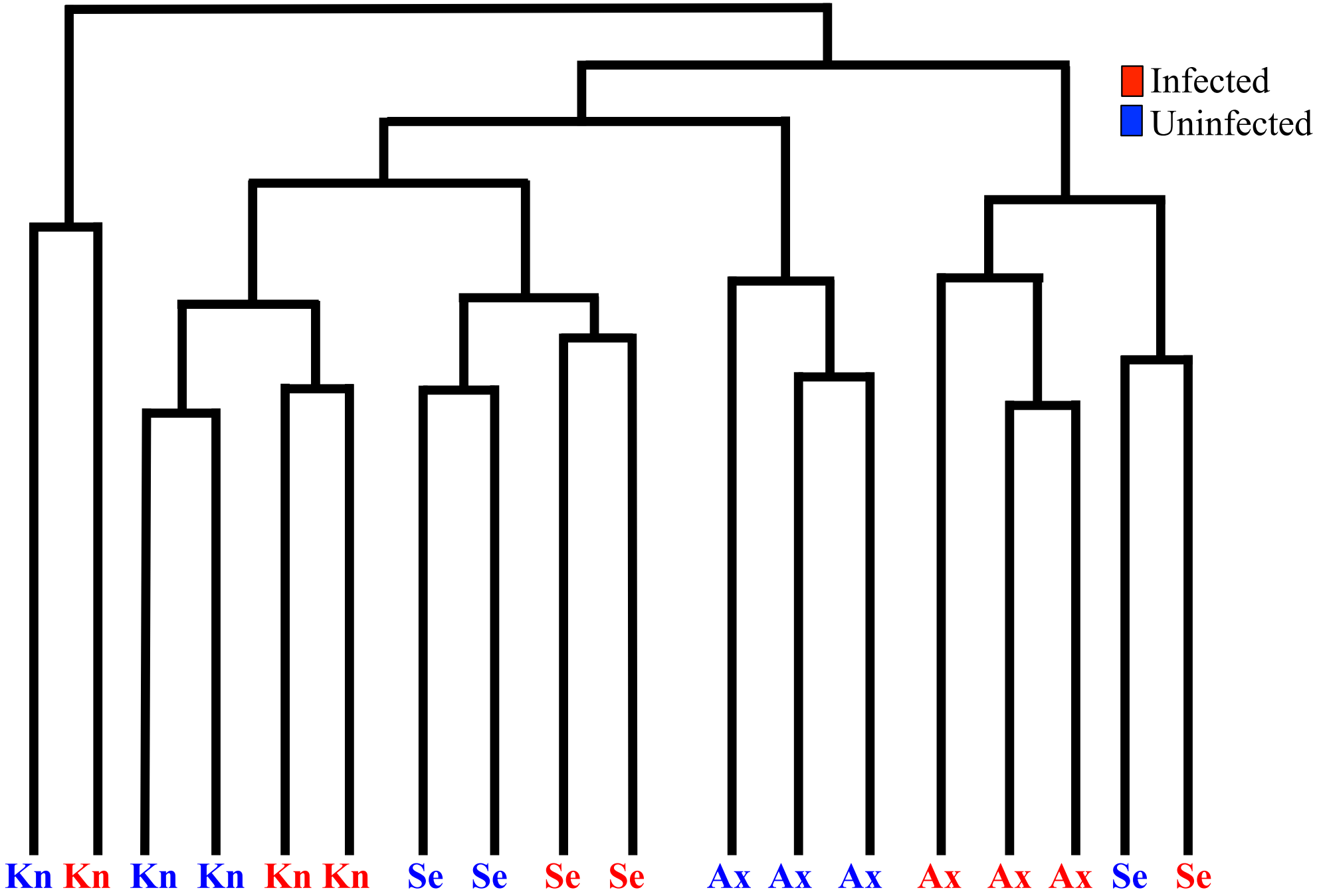


Fig. S3. Dendrogram (generated with cummeRbund) depicting how expression profile is influenced by lake of origin and infection status. Samples are clustered based on their expression profile using the Jensen-Shannon distance. Samples that have more similar expression profiles are grouped more closely together. These results recapitulate and strengthen those depicted in the MDS plot (Fig. 4).