PTC2 region genotypes counteract *Biomphalaria glabrata* population differences between M-line and BS90 in resistance to infection by *Schistosoma mansoni*

Michael S Blouin Corresp., 1 , **Stephanie R Bollmann** [2] , **Jacob A Tennessen** [3]

1 Integrative Biology, Oregon State University, Corvallis, Oregon, United States of America

2 Integrative Biology, Oregon State University, Corvalis, Oregon, United States of America

3 Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America

Corresponding Author: Michael S Blouin
Email address: blouinm@science.oregonstate.edu

Background. *Biomphalaria glabrata* is a snail intermediate host for *Schistosoma mansoni,* a trematode responsible for human schistosomiasis. BS90 is one of the most well studied strains of *B. glabrata* owing to its high resistance to infection by most strains of *S. mansoni*. An F2 mapping study from 1999 identified two RAPD markers that associated with what appeared to be single-locus, dominant resistance by the BS90 population relative to the susceptible M-line population. One marker cannot be mapped, but the other, *OPM-04*, maps to within 5 Mb of *PTC2*, a region we recently showed has a very large effect on resistance within another snail population challenged by the same strain of parasite (PR1). Here we tested the hypothesis that the *PTC2* region contains the causal gene/s that explain the iconic resistance of BS90 snails.

Methods. We used marker-assisted backcrossing to drive the BS90 version of the PTC2 region (+/- ~ 1 Mb on either side) into an M-line (susceptible strain) genetic background, and the M-line version into a BS90 genetic background. We challenged the offspring with PR1-strain schistosomes and tested for effects of allelic variation in the *PTC2* region in a common genetic background.

Results. Relative to M-line haplotypes, the BS90 haplotype actually confers enhanced susceptibility. So we reject our original hypothesis. One possible explanation for our result was that the causal gene linked to *OPM-04* is near, but not in the *PTC2* block that we introgressed into each line. So we used an F2 cross to independently test the effects of the *PTC2* and *OPM-04* regions in a randomized genetic background. We confirmed that the BS90 haplotype confers increased susceptibility, and we see a similar, although non-significant effect at *OPM-04*. We discuss possible reasons why our results differed so dramatically from those of the 1999 study. We also present Pacbio assemblies of the *PTC2* and flanking region in BS90 and M-line, compare with previously published *PTC2* haplotypes, and discuss candidate genes that might be behind the enhanced susceptibility of the BS90 haplotype.

5   *PTC2* region genotypes counteract *Biomphalaria glabrata* population differences between M-line and

6   BS90 in resistance to infection by *Schistosoma mansoni*

11   Blouin, Michael S.[1], Stephanie R. Bollmann,[1] Jacob A. Tennessen[2]

16   [1]Department of Integrative Biology, Oregon State University, Corvallis, Oregon, USA

17   [2]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health,

18   Boston, Massachusetts, USA

22   Corresponding Author:

23   Michael Blouin

24   Cordley 3029, Dept. Integrative Biology, Oregon State University, Corvallis, OR 97330 USA

25   Email address: blouinm@oregonstate.edu

Abstract.

29 **Background**. *Biomphalaria glabrata* is a snail intermediate host for *Schistosoma mansoni,* a

30 trematode responsible for human schistosomiasis. BS90 is one of the most well studied strains

31 of *B. glabrata* owing to its high resistance to infection by most strains of *S. mansoni*. An F2

32 mapping study from 1999 identified two RAPD markers that associated with what appeared to be

33 single-locus, dominant resistance by the BS90 population relative to the susceptible M-line

34 population. One marker cannot be mapped, but the other, *OPM-04,* maps to within 5 Mb of

35 *PTC2*, a region we recently showed has a very large effect on resistance within another snail

36 population challenged by the same strain of parasite (PR1). Here we tested the hypothesis that

37 the *PTC2* region contains the causal gene/s that explain the iconic resistance of BS90 snails.

38

39 **Methods**. We used marker-assisted backcrossing to drive the BS90 version of the PTC2 region

40 (+/- ~ 1 Mb on either side) into an M-line (susceptible strain) genetic background, and the M-line

41 version into a BS90 genetic background. We challenged the offspring with PR1-strain

42 schistosomes and tested for effects of allelic variation in the *PTC2* region in a common genetic

43 background.

44

45 **Results**. Relative to M-line haplotypes, the BS90 haplotype actually confers enhanced

46 susceptibility. So we reject our original hypothesis. One possible explanation for our result was

47 that the causal gene linked to *OPM-04* is near, but not in the *PTC2* block that we introgressed

48 into each line. So we used an F2 cross to independently test the effects of the *PTC2* and *OPM-*

49 *04* regions in a randomized genetic background. We confirmed that the BS90 haplotype confers

50 increased susceptibility, and we see a similar, although non-significant effect at *OPM-04*. We

51 discuss possible reasons why our results differed so dramatically from those of the 1999 study.

52 We also present PacBio assemblies of the *PTC2* and flanking region in BS90 and M-line,

53 compare with previously published *PTC2* haplotypes, and discuss candidate genes that might be

54 behind the enhanced susceptibility of the BS90 haplotype.

55

56

57

Introduction

Schistosomiasis is a trematode parasitic disease of humans that affects over 200 million people in tropical regions worldwide, causing a chronic, debilitating condition that kills over 200,000 people annually (WHO, 2022). The life cycle of schistosomes requires passage through an aquatic snail intermediate host. Humans are infected by spending time in water inhabited by infected snails. *Biomphalaria glabrata* is a New World snail that is the most well-studied intermediate host of *S. mansoni*. Resistance to *S. mansoni* is highly heritable and easily selected for in *B. glabrata* (Richards and Shade, 1987; Richards et al. 1992; Webster et al. 2004; Tennessen, Theron et al. 2015). So there is substantial interest in using resistance alleles to genetically alter natural snail populations to make them less able to transmit the parasite (Marques et al., 2014; Famakinde, 2018; Maier et al., 2019; Grewelle et al., 2021), as is being developed for mosquitos (Pham et al., 2019; Adolfi et al., 2020; Carballar-Lejarazú et al., 2020). However, the key genes and molecular pathways underlying this genetic variation for resistance in snails are not yet well understood.

Genome-wide mapping studies using various strains of *B. glabrata* and *S. mansoni* have revealed five genomic regions in which allelic variation controls resistance to *S. mansoni* (Knight et al., 1999; Goodall et al., 2006; Tennessen, Theron et al., 2015; Tennessen, Bonner et al. 2015; Tennessen et al., 2020). Here we use the terms "susceptibility" and "resistance" to mean the percentage of snails in a population that do, or do not, get infected after individuals are challenged with a fixed number of schistosome miracidia. We also use the terms "strain" and "population" interchangeably, acknowledging that most of the "strains" referred to in the literature are genetically variable.

The region showing the strongest effect on resistance to date is the *PTC2* (Polymorphic transmembrane complex 2) region. This region was identified via a genome-wide association study (GWAS) using outbred 1316R snails challenged by PR1 schistosomes (Tennessen et al., 2020). In 1316R snails the *PTC2* is a 300 to 450 kb region that contains 11 genes, 9 of which code for single-pass transmembrane proteins. Haplotypes for the *PTC2* region are highly divergent in both structure and sequence. Haplotypes vary by multi-kb indels and some genes are found on only one or two of the three haplotypes. The enrichment for transmembrane proteins in this region is similar to what was found in the *GRC* (= *PTC1*) region (Tennessen, Theron et al., 2015), which suggests a model in which

88 molecular recognition (either of the parasite by the host, or vice versa) plays an important role in snail-

89 schistosome compatibility polymorphisms (e.g. Theron et al., 2008).

90

91 BS90 is a Brazilian strain of *B. glabrata* that is highly resistant to most strains of *S. mansoni*. Thus, it

92 is one of the most-well-studied strains of *B. glabrata*, the subject of dozens, if not hundreds, of

93 functional studies on its resistance over the last 50 years (e.g. Zhang et al., 2008; Yoshino et al., 2012;

94 Ittiprasert and Knight 2012; Pila, Gordy et al., 2016; Pila, Tarrabain et al., 2016). Therefore, there is

95 substantial interest in finding the loci that make BS90 so much more resistant than other strains of *B.*

96 *glabrata*. Knight et al. (1999) performed an F2 QTL mapping cross to study the difference in

97 resistance between BS90 and M-line strain snails to the SmPR1 and NMRI strains of *S. mansoni*. M-

98 line is highly susceptible to SmPR1 and NMRI *S. mansoni*, while BS90 is 100% resistant to them, and

99 the difference between snail strains appeared to segregate as a simple Mendelian trait with the BS90

100 allele dominant. Knight et al. used either SmPR1 or NMRI strain *S. mansoni* in their experiments, as

101 the two strains behaved identically towards M-line and BS90 (M. Knight, pers. comm. June 2022).

102 Knight et al. scored RAPD markers in the F2s and reported the sequences of two bands that associated

103 with resistance (RAPD is an older DNA fingerprinting technique that relies on low-stringency PCR

104 amplification using single, arbitrary, 10mer primers). After the publication of the *B. glabrata* genome

105 it became clear that one of their bands (OPZ-11) is embedded in repetitive sequence that is dispersed

106 throughout the genome, and so could not be localized. However, the other band (locus OPM-04) maps

107 to a single location in the *B. glabrata* genome (Tennessen et al., 2020).

108

109 In our PacBio assemblies of *B. glabrata* genomes, *OPM-04* is only 5 Mb from the *PTC2* region

110 (Tennessen et al., 2020). Furthermore, from our linkage map (Tennessen et al., 2017) we estimated

111 that *OPM-4* is about 20 cM from the PTC2 region, which is similar to the 17 cM recombination

112 distance one estimates OPM-04 to be from the causal gene, based on Knight et al.'s published

113 infection data (Tennessen et al., 2020). So we hypothesized that the *PTC2* region contains the actual

114 causal gene/s linked to the *OPM-04* marker that makes BS90 so much more resistant than M-line.

115 Here we report a test of that hypothesis, in which we used marker-assisted backcrossing to drive the

116 BS90 version of the PTC2 region (+/- ~ 1 Mb on either side) into an M-line genetic background, and

117 the M-line version into a BS90 genetic background. We find that segregation of the BS90 and M-line

118 alleles in a common genetic background has a significant effect on resistance to PR1 schistosomes, but

119   that the BS90 allele confers increased susceptibility. So this experiment rejects the hypothesis that the

120   PTC2 region explains the difference in susceptibility between BS90 and M-line.   Furthermore, using

121   an F2 cross we could not replicate Knight et al.'s (1999) result that a BS90 allele at *OPM-04*

122   segregates with resistance.

123

Materials & Methods

125   *Snail strains*

126   The original M-line was reportedly created in the 1950s by crossing a pigmented Puerto Rican strain

127   that was susceptible to Puerto Rican schistosomes, with an albino, resistant Brazilian strain, and then

128   inbreeding with selection for albinism and susceptibility (Newton, 1955; Richards and Merritt, 1972).

129   BS90 was reportedly isolated in Salvador, Bahia, Brazil, in the 1960s, and has been in labs in the US

130   since the 1990s (Paraense and Correa, 1963, cited in Knight et al., 1999; C. Bayne, 2006, pers.

131   comm.). Both have been maintained by the Biomedical Research Institute (BRI) schistosomiasis

132   resource center (https://www.afbr-bri.org/schistosomiasis/) and in dozens of labs around the world for

133   decades. We obtained BS90 in 2015 from Timothy Anderson's lab at the Texas Biomedical Research

134   Institute (originally from BRI), and M-line snails in 2018 from the BRI (their "T0" population).   For

135   our crosses we used the outbred M-line population and an inbred line of BS90 that we generated via

136   two generations of selfing (line FSS5). Snails were housed in 7.5 liter plastic tanks containing

137   artificial spring water (Ulmer, 1970).

138

139   *Schistosome challenges*

140   We challenged 6 to 9 mm diameter snails individually in 24-well plates as follows: 15 miracidia of

141   PR1-strain *S. mansoni* were added to ~1 ml water in each well of a 24-well plate. Then a single snail

142   was put into each well. Snails were left in the plates with miracidia for two hours, and were then

143   moved to darkened tubs of ~24 snails per tub. Snails were kept in the darkened tubs for four weeks,

144   and then were checked every week or two for shedding for six more weeks. Snails that did not shed

145   within the 10 week window were classified as non-infected. Individuals that died before shedding

146   were not included in the analysis. BS90 is always 100% resistant to PR1 (i.e. no BS90 snails ever got

147   infected), while about 60% of our M-line snails get infected under these challenge conditions. In all

148   experimental challenges for this study we always included at least 12 pure BS90 and 12 pure M-line

149   snails as controls for schistosome infectivity.   The Oregon State University Institutional Animal Care

150 and Use Committee provided approval for this research under Animal Care and Use Protocols 2021-
151 0213 and 5115.

152

153 *PacBio sequencing and characterization of the PTC2 region in BS90 and M-line*
154 We used an inbred line of BS90 (2 generations of selfing) and an outbred population of M-line for our
155 crosses. We obtained PacBio genomic assemblies from a snail from the inbred BS90 line and from an
156 M-line snail. The BS90 and M-line samples were run on a Sequel II machine and assembled using
157 Flye (Kolmogorov et al., 2019). We generated assemblies with high contiguity for both BS90 (2534
158 fragments, length = 843.8 Mb, N50 = 2.3 Mb, largest scaffold = 15.5 Mb) and M-line (1834
159 fragments, length = 852.5 Mb, N50 = 15.1 Mb, largest scaffold = 38.5 Mb). In both genomes, *PTC2*
160 occurs on assembled fragments over 4 Mb long, more than 1 Mb from either end, allowing for detailed
161 assessment of the surrounding genomic region. We used alignments of the BS90 and M-line sequences
162 to find regions of synteny around the *PTC2* region in which we could place PCR markers for the
163 marker-assisted backcrossing (Fig 1). We used GENSCAN (Burge and Karlin, 1997), HMMER (Finn
164 et al. 2011), DeepTMHMM (Hallgren et al. 2022), Vectorbase (Giraldo-Calderón et al. 2015), and
165 BLASTP to annotate coding regions in these assemblies, supplemented by our previous genomic and
166 transcriptomic data from the 1316R population of *B. glabrata* (Tennessen et al. 2020). Assemblies
167 have been made available through NCBI Genome, BioProject Accession PRJNA847915.

168

169 *Backcross experiment*
170 We used PCR to score five markers in and around the *PTC2* region (Fig. 1). For ease of scoring, we
171 designed primers to target gel-scorable indels that differed between our M-line and BS90 parental
172 strains (Table S1). To drive the BS90 *PTC2* region into an M-line genetic background we created F1s,
173 and then backcrossed them to pure M-line ("B into M" backcrossing). We genotyped backcross
174 offspring until we found recombinants inside one of the outer two markers, and then used those
175 offspring in repeated rounds of backcrossing and genotyping until we found individuals that were
176 homozygous M-line at the two outer markers and heterozygous at the inner three markers (i.e.
177 genotype MHHHM at the five loci, where M = homozygous for M-line allele and H = heterozygous;
178 Fig. 1). This took 3 or 4 rounds of backcrossing per tank. Individual snails with the desired double
179 recombination event (MHHHM) were backcrossed one final time to further remove background BS90
180 genome and to produce a large number of siblings which all had the MHHHM genotype. We then let

181   tanks of MHHHM siblings breed randomly and expand to create populations containing MMMMM,

182   MHHHM and MBBBM individuals on a randomized genetic background for challenge testing. Each

183   tank was started from an independent double recombinant event. We created and tested four

184   independent "B into M" populations (i.e. started by MHHHM individuals), named M8.2, M8-2.2,

185   M13.2 and M17. We challenged and genotyped N = 159, 89, 85 and 68 surviving snails from each of

186   these populations, respectively.

188   We used the same genotyping and backcross approach to drive the M-line *PTC2* region into a BS90

189   genetic background ("M into B" backcross population). We were able to find only a single BHHHB

190   individual (where B = homozygous for the BS90 allele). We used that snail to create a single "M into

191   B" population (population B7.2), following the same protocol as with the "B into M" populations.

192   B7.2 produced only 11 infected individuals out of 116 challenged survivors (9.2%).

194   *F2 Intercross population to test regional effect*

195   We also crossed BS90 x M-line to create an F2 cross population that we allowed to breed randomly for

196   almost two years while we conducted the backcrossing experiments (this population was created as a

197   backup mapping population in case we rejected the *PTC2* hypothesis). We pooled F1s from two

198   individual BS90 x M-line individual snail crosses, and then used their offspring to found the F2

199   breeding population. Our lab was abruptly shut down during the covid pandemic, during which time

200   we were able to maintain the F2 breeding population. However, we lost the tubs that contained the

201   original founding parents, so we don't have their DNA. The BS90 parents were all from an inbred

202   line, so are likely to be similar to the source inbred line, which we retain and for which we have a

203   PacBio assembly. As with the backcross populations, M-line parents were from our outbred

204   population. So there could be genetic variation within the M-line contribution to the F2 population.

206   Seven months after setting up the F2 population, we started challenging ~two dozen snails every few

207   months until we had accumulated > 300 phenotyped individuals. We challenged a total of 424

208   individuals, of which 333 survived and 62 (19%) became infected. We successfully genotyped 329 of

209   these 333 individuals at the 5 loci that span the backcross region. We chose 55 infected and 55 non-

210   infected individuals as cases and controls, and additionally genotyped them via PCR at *OPM-04*

211   (primers in Table S1) in order to compare the effects of allelic variation at *OPM-04* vs *PTC2*. The

212 *OPM-04* amplicons were scored via Sanger sequencing. In the full sample of 333 individuals we

213 observed evidence of null alleles at marker loci *0, dn1* and *dn2*, while markers *up1* and *up2* showed

214 Hardy-Weinberg Equilibrium. So we analyzed genotypes at marker *up1* as a proxy for *PTC2* genotype

215 (only ~1 Mb away, so should be in high linkage disequilibrium with marker *0*). We noticed no

216 evidence for null alleles in the backcross populations, but that is expected because each generation we

217 used individuals that were successfully typed at all five loci.

218

219 *Statistical analysis*

220 The various populations created, and samples used, for this study are summarized for clarity in Fig.

221 S1. The effect of segregation for BS90 vs M-line alleles on infection status in the backcross and F2

222 populations was analyzed via binary logistic regression in Systat 13.2. For the four "B into M"

223 backcross populations, we tested main effects of genotype and population, and the genotype x

224 population interaction. We tested additive and completely dominant modes of inheritance. We

225 observed too few infected snails from the "M into B" population to have any power for a statistical

226 test, so we simply show the data from B7.2 and note that the general pattern is consistent with what we

227 observed in the "B into M" populations.

228

229 We tested the effect of *PTC2* genotypes using marker *0* in the backcross populations, and using marker

230 *up1* in the F2 population (to avoid the null allele at marker *0*). For locus *OPM-04*, we observed two

231 sequences from the M-line population segregating in the F2s, both of which were distinct from the

232 BS90 sequence (Fig. S2). So we analyzed those data as bi-allelic, *B* vs *M* (where *M* = either of the two

233 M-line alleles).

234

Results

236 *Effect of PTC2 in mostly M-line genetic backgrounds*

237 The number of infected and not-infected individuals per *PTC2* genotype in each "B into M" backcross

238 population is shown in Figure 2. The additive model showed only a slightly better fit than an M-

239 allele-dominant model (AIC = 458.8 vs 459.6). There were strong main effects of both genotype and

240 population on infection status (Table S2, S3). The genotype-by-population interactions were not

241 significant, despite the visual appearance of a different genotypic effect in population 17 in Fig. 2.

242 Under an additive model, each additional BS90 allele roughly doubles the odds of infection (odds ratio

243 ± standard error = 2.2 ± 0.78; P = 0.025). The estimated effect is similar under an M-allele dominant

244 model (odds ratio = 3.8 ± 1.9 for the *BB* genotype relative to *M_* = *MM* or *BM*; P = 0.009).

245

246 One unexpected result is that we see a significant heterozygote excess at the *PTC2* in population M8.2,

247 and a trend of heterozygote excess in the other three "B into M" populations (F_{IS} at locus *0* = -0.18, -

248 0.12, -0.06 and -0.21 in populations M8.2, M8-2.2, M13.2 and M17, respectively). Combining all four

249 populations together, observed heterozygotes (N = 228) significantly exceed expectation (χ^2 test, P =

250 0.003), whether expected heterozygosity is calculated from each population separately (expected N =

251 198.7) or from all samples pooled (expected N = 199.7). We see similar negative F_{IS} at all three of the

252 segregating backcross-region loci (*up1, 0* and *dn 1*) in these populations, so the negative F_{IS} seems

253 unlikely to result from technical artifacts.

254

255 *Effect of PTC2 in a mostly BS90 genetic background*

256 We see infection in all three genotypes, and the trend is consistent with what we observed in the four

257 "B into M" populations (fractions infected were 3/33 = 0.11 *BB*, 7/64 = 0.11 *BM*, and 1/19 = 0.06

258 *MM*). Population B7.2 also showed a trend of heterozygote excess at the backcross region loci (F_{IS} = -

259 0.12 at all three loci), although the deviation from Hardy Weinberg Equilibrium is not significant.

260

261 *Effect of PTC2 in the F2-cross genetic background, and comparison with OPM-04*

262 We examined effects of the *PTC2* and *OPM-04* regions in the F2 population in order to test whether

263 the locus linked to *OPM-04* in Knight et al. (1999) was still nearby, even if not within the region we

264 introgressed into our backcross lines. As in the backcross populations, the BS90 allele conferred

265 significantly enhanced susceptibility in the full sample from the F2 population (N= 329 successfully

266 genotyped) (Table S2, S3, Fig. 3). The additive and M-allele-dominant models were equally

267 supported. Odds ratios for increased susceptibility were 1.5 (± 0.32) per *B* allele under an additive

268 model (P= 0.041) and 1.8 (± 0.54) for *BB* relative to *M_* under an M-allele dominant model (P =

269 0.044).

270

271 The subsample of 55 cases and 55 controls also showed significantly enhanced susceptibility caused

272 by the *PTC2 B* allele. Odds ratios = 2.0 (± 0.61; P = 0.029) and 3.3 (± 1.49; P = 0.007) for the

273 genotype effect under additive and M-allele dominant models, respectively (Table S2, S3). There was

274   a similar pattern at *OPM-04* (Fig. 4), although the genotype effect was not significant (P = 0.286 and

275   0.380 for additive and M-allele dominant models, respectively; Table S2, S3). So unlike our original

276   expectations from Knight et al. (1999), *OPM-04* does not appear to be anywhere near a locus at which

277   the BS90 allele confers dominant resistance. Instead, it shows a pattern similar to that of the closely-

278   linked *PTC2* region.

279

280   As mentioned previously, there were obvious null alleles at loci *0* to *dn2*. But loci *up2* and *up1* were in

281   Hardy-Weinberg Equilibrium (F_{IS} = -0.03), so we do not see the pattern of heterozygote excess that

282   was apparent in the backcross populations.

283

284   *PacBio assemblies*

285   In the 1316R population there are three *PTC2*-region haplotypes that are ordered from most to least

286   protective as *R* > *S2* > *S1* (Tennessen et al. 2020). Our M-line haplotype is over 99% similar to the

287   previously described *S2* haplotype, with a total of only six nonsynonymous differences between M-

288   line and *S2* across all 11 genes in the region. Our BS90 haplotype is distinct from other known

289   haplotypes; a portion of it is similar to *S2*/M-line, while the remainder is closer to *S1* and is extremely

290   divergent from *S2*/M-line (Fig. 5, 6). At the amino acid level, *PTC2* genes 1-5 are 95% identical

291   between BS90 and *S2*/M-line, while the remaining genes are either only present in BS90 (gene 7), only

292   present in M-line (genes 6 and 8), or in the case of gene 9 (*BGLB018951*) present in both but over

293   50% divergent in sequence between them. Thus, this divergent portion of *PTC2* seems most likely to

294   explain the observed phenotypic difference between BS90 and M-line.

295

296   Recall, however, that the ~300 kb *PTC2* region is embedded within a larger region (~850 kb in BS90

297   and ~1300 kb in M-line) that shows almost no sequence similarity between our M-line and BS90

298   assemblies (Fig. 6). Thus, it is possible that the causal gene/s lie in the flanking regions, which we

299   also introgressed. Both haplotypes possess unique genes in this zone, although as with many snail

300   genes, little is known about their function and most appear to be transposable elements. Notably, three

301   AIG genes occur in this region in M-line but not BS90 (in possession of an AIG1 domain; homologs

302   of each other and of *BGLB016854*, *BGLB012732*, and *BGLB004870*). AIG genes are suspected to

303   have immune function in *Biomphalaria* (Lu et al., 2020). We have confirmed with RNA-Seq data

304   from haplotype *S2* (Accession SRX8534561) that all three AIG1 genes are expressed. Farther out from

305   *PTC2*, our BS90 and M-line assemblies are again similar in sequence and so seem less likely to

306   contain the causal genes, though this remains a possibility. For example, a toll-like receptor gene

307   *BGLB025483* occurs near locus *dn1* in both populations and has 24 amino acid differences between

308   the haplotypes. Pila, Tarrabain et al. (2016) showed that knocking down expression of a different Toll-

309   like receptor makes BS90 more susceptible, so this is an interesting class of proteins.

310

311   One interesting observation from our BS90 and M-line PacBio assemblies is that the 1.2 kb *OPM-04*

312   amplicon sequence published by Knight et al. (genbank AF078109) is almost the same in our BS90

313   assembly, but is interrupted by a ~1.7 kb insertion in our M-line assembly (Fig. S2). So if Knight et

314   al.'s M-line and BS90 also differed at this indel, then that might explain why under a low-stringency

315   RAPD-PCR reaction they got amplification from BS90 but not from M-line chromosomes, resulting in

316   a dominant marker for BS90.

317

318   **Discussion**

319   We have conclusively rejected the hypothesis that the *PTC2* region explains the difference between

320   our M-line and BS90 populations in resistance to PR1-strain *S. mansoni*. Nor could we replicate

321   Knight et al.'s (1999) result that the *OPM-04* locus was linked to a dominant resistance from BS90.

322   Furthermore, in a common genetic background, the BS90 allele actually conferred increased

323   susceptibility relative to the M-line alleles. This is the opposite of what was predicted, given our BS90

324   remains 0% susceptible to PR1 schistosomes, while our M-line is 60% susceptible.

325

326   *Why were we unable to replicate Knight et al.'s (1999) result?*

327   We were surprised to see no effect of the OPM-04 region in our BS90 x M-line F2 population.

328   Possible explanations for this failure to replicate Knight et al.'s result include: (1) the strains of

329   "BS90", "M-line" or schistosome we used differ from theirs, and (2) we were not measuring the same

330   phenotype.

331

332   (1) BS90 and M-line have been in culture in many labs for decades, so there is no guarantee that our

333   lines are genetically the same as the those used by Knight et al. (1999) over 20 years ago. BS90 is

334   phenotypically consistent in that it remains 0% susceptible to PR1 schistosomes. On the other hand,

335   Knight et al. (1999) reported that their M-line was 100% susceptible to PR1, whereas under our

336 experimental conditions M-line was only ~60% susceptible. Thus, our outbred M-line population may
337 contain variation for resistance. Perhaps their M-line was more isogenic and carried a susceptibility
338 allele near *OPM-04* that is absent or in low frequency in our M-line. Another possibility is that effects
339 of genetic background change the penetrance of whatever allele was linked to *OPM-04* in their M-line.
340 It is also possible that the SmPR1 strain of schistosome we used differs in some way from the strains
341 they used.

342

343 (2) Another plausible explanation for the difference in results between the two studies is that we were
344 not measuring the same phenotype. In Knight et al.'s (1999) study they report phenotype ratios in
345 BS90 x M-line F1s and F2s that are consistent with single-locus Mendelian inheritance, and in which
346 the BS90 allele is dominant and confers complete resistance (i.e. 0% infected F1s, ~25% infected F2s).
347 Older literature on crosses among strains that differed in resistance to PR1 reported data consistent
348 with single-locus inheritance when using mature snails, but data more consistent with polygenic
349 inheritance when using juvenile snails (e.g. Richards, 1984). Thus, it is possible that our experimental
350 conditions (e.g. snail size or maturity, number of miracidia per challenge, or some other environmental
351 factor) reduced the importance of whatever highly penetrant genotype was linked to *OPM-04* in
352 Knight et al.'s study. For example, although both studies used similar-sized snails, Knight et al.
353 (1999) challenged each snail with 100 miracidia, whereas we challenged with only 15 miracidia
354 (because higher numbers caused excess mortality in our hands).

355

356 There is no obvious reason to doubt the data in Knight et al. (1999). Plus, it seems a remarkable
357 coincidence that in a 0.9 Gb genome of 18 chromosomes, such an important region as *PTC2* would lie
358 only 5 Mb, and almost exactly the expected number of cM from *OPM-04*, yet not contain the causal
359 gene/s (Tennessen et al., 2020). Therefore, we suggest that if one could reproduce the highly-
360 penetrant, dominant BS90 phenotype observed by Knight et al. (1999) and previous workers, then it
361 would be worth re-testing whether the OPM-04 region shows an association, and whether the causal
362 gene/s linked to that marker lie in the *PTC2* region.

363

364 *Genetic background effects*
365 The sizable population effects among the four "B into M" populations were stronger than the genotype
366 effect (Fig. 2; Table S3), and presumably reflect their different genetic backgrounds. This shows that

367   other loci have a very large effect on resistance in this system. Each population was started using

368   different M-line individuals from the outbred M-line, and each must retain different residual segments

369   of donor line (BS90) genome that is still segregating. So there is plenty of opportunity for these

370   populations to differ in genetic background, and we don't know to what extent that results from

371   variation in the base M-line population or from different retained segments of BS90 chromosomes.

372   The population-by-genotype interactions were not significant, although the trend towards a weaker

373   genotypic effect in population M17 was intriguing (Fig. 2). The effect of SOD1-region alleles on

374   resistance appears to depend on genetic background (Bonner et al., 2012), so similar epistatic effects

375   would not be surprising.

376

377   *Negative F_{IS} in the PTC2 region*

378   It was intriguing that F_{IS} in the *PTC2* region was negative in all of the backcross populations, and

379   significantly so in one of them, and in the set of all populations combined. The mortality rate after

380   parasite challenge in these populations ranged from 6% to 31%, and if these deaths were primarily *BB*

381   and/or *MM* homozygotes, that could entirely explain the observed heterozygote excess. However,

382   observed mortality was not correlated with F_{IS} (for example, population 13.2 had the highest mortality

383   but the lowest heterozygote excess) and heterozygotes did not show any overdominant effect on odds

384   of infection (Fig. 2, 3). So if there is overdominance for survival, it is not obviously owing to the

385   schistosome challenges. Unfortunately we did not keep the backcross populations at the end of the

386   experiment, so we cannot go back and test for Hardy-Weinberg equilibrium in the unchallenged

387   populations. A slightly negative F_{IS} can also result when a small effective number of breeders (N_b)

388   contributes to a sample, but to get F_{IS} in the observed range of -0.15-0.20 would require $N_b \sim 2$-3

389   individuals (Waples, 2015). Such a small N_b is not out of the question, but seems low. That there

390   might be overdominance for survival in a region of such obvious immune relevance is a hypothesis

391   worthy of further exploration, as this would be one mechanism that could maintain genetic diversity in

392   the region.

393

394   *Additional insights into the PTC2 region*

395   Allelic variation in the *PTC2* region has the largest known effect on resistance identified to date

396   (Tennessen et al., 2020). The extreme structural and non-synonymous sequence variation in this

397   region, and its enrichment for membrane-spanning protein genes, strongly suggests host-parasite

398   interacting molecules under selection. Of course, schistosomes are an introduced species in the New

399   World, so this region no doubt evolved with other pathogens. It will be important to determine which

400   loci are involved in interactions with schistosomes, and if multiple loci are involved, how they interact.

401   The 2- to 4-fold effect of *BB* versus *MM* observed here was weaker than previously reported for *PTC2*

402   (Tennessen et al. 2020), but comparable to the effect of other important *B. glabrata* resistance loci like

403   *sod1* or *RADres* (Tennessen, Bonner et al 2015). Such effects are real and likely to be biologically

404   meaningful, but they are easily missed. Here we benefitted from the use of hundreds of snails. A

405   power analysis shows that if we had used half the sample size in this study, we likely would not have

406   inferred a significant genotype effect. Thus lower-powered studies might miss these more subtle

407   genotype-phenotype correlations.

408   Although our addition of two new assemblies of the region revealed new structural variants and

409   additional sequence variation, the two new haplotypes closely resemble the other known haplotypes or

410   recombinants of them. Comparing the two new haplotypes to the three previously described

411   haplotypes allowed us to make some *a priori* predictions about which loci are most likely to be

412   involved. For example, the M-line haplotype closely resembles *S2* while the BS90 haplotype is closer

413   to *S1*, and thus the 7-fold decrease in infection odds conferred by *S2S2* relative to *S1S1* (Tennessen et

414   al. 2020) could have the same genetic basis as the effect reported here, perhaps due to one of genes 6

415   through 9 or an AIG gene. We are currently attempting functional knockdowns of these genes to

416   provide definitive tests.

417

418   Conclusions

419   The most striking result of this work is that not only did the block containing *OPM-04* and *PTC2* not

420   contain the dominant-resistance allele predicted by Knight et al. (1999), the BS90 version actually

421   conferred susceptibility. This antagonistic effect must be completely counteracted by other loci that

422   remain to be identified, as our BS90 remains 100% resistant to PR1 schistosomes, and M-line highly

423   susceptible. It will be extremely interesting to find those additional loci and to finally understand how

424   they interact with whatever locus Knight et al. (1999) mapped over 20 years ago. Deciphering this

425   complex, polygenic basis for resistance will be facilitated by the emerging plethora of *Biomphalaria*

426   genome assemblies, including those presented here and others (e.g. Nong et al. 2022). In turn, these

427   discoveries will enable new options for vector-based disease control.

428

429   Acknowledgements

430 PacBio library preparation and/or PacBio sequencing was done by the Center for Quantitative
431 Life Sciences Core Lab at Oregon State University, and the University of Oregon Genomics and
432 Cell Characterization Core Lab. M-line snails and SmPR1 schistosomes were provided by the
433 NIAID Schistosomiasis Resource Center of the Biomedical Research Institute (Rockville, MD)
434 through NIH-NIAID Contract HHSN272201700014I. Ed Davis helped with use of Flye. We
435 thank Chloe Hanson, Maelyn Estrada, Marisa Bennett, Isabelle Ray, Abigail LaVerdure and
436 Havana Soler for help with snail husbandry.

437

References

439 Adolfi A, Gantz VM, Jasinskiene N, Lee H-F, Hwang K, Terradas G, Bulger EA, Ramaiah A,
440 Bennett JB, Emerson JJ, Marshall JM, Bier E, James AA. 2020. Efficient population
441 modification gene-drive rescue system in the malaria mosquito Anopheles stephensi.
442 *Nature Communications* 11:5553. DOI: 10.1038/s41467-020-19426-0.

443 Bonner KM, Bayne CJ, Larson MK, Blouin MS. 2012. Effects of Cu/Zn Superoxide Dismutase
444 (sod1) Genotype and Genetic Background on Growth, Reproduction and Defense in
445 Biomphalaria glabrata. *PLOS Neglected Tropical Diseases* 6:e1701. DOI:
446 10.1371/journal.pntd.0001701.

447 Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J
448 Mol Biol* 268:78–94. DOI: 10.1006/jmbi.1997.0951.

449 Carballar-Lejarazú R, Ogaugwu C, Tushar T, Kelsey A, Pham TB, Murphy J, Schmidt H, Lee
450 Y, Lanzaro GC, James AA. 2020. Next-generation gene drive for population modification
451 of the malaria vector mosquito, Anopheles gambiae. *Proceedings of the National Academy
452 of Sciences* 117:22805–22814. DOI: 10.1073/pnas.2010214117.

453 Cohen LM, Neimark H, Eveland LK. 1980. Schistosoma mansoni: Response of Cercariae to a
454 Thermal Gradient. *The Journal of Parasitology* 66:362–364. DOI: 10.2307/3280843.

455 Famakinde DO. 2018. Treading the Path towards Genetic Control of Snail Resistance to
456 Schistosome Infection. *Tropical Medicine and Infectious Disease* 3. DOI:
457 10.3390/tropicalmed3030086.

458 Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity
459 searching. *Nucleic Acids Research* 39:W29–W37. DOI: 10.1093/nar/gkr367.

460 Giraldo-Calderón GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, Ho N,
461 Gesing S, VectorBase Consortium, Madey G, Collins FH, Lawson D. 2015. VectorBase: an

462   updated bioinformatics resource for invertebrate vectors and other organisms related with

463   human diseases. *Nucleic Acids Research* 43:D707-713. DOI: 10.1093/nar/gku1117.

464 Goodall CP, Bender RC, Brooks JK, Bayne CJ. 2006. Biomphalaria glabrata cytosolic

465   copper/zinc superoxide dismutase (SOD1) gene: Association of SOD1 alleles with

466   resistance/susceptibility to Schistosoma mansoni. *Molecular and Biochemical Parasitology*

467   147:207–210. DOI: 10.1016/j.molbiopara.2006.02.009.

468 Grewelle RE, Perez-Saez J, Tycko J, Namigai EKO, Rickards CG, Leo GAD. 2021. *Modeling*

469   *the efficacy of CRISPR gene drive for schistosomiasis control.* DOI:

470   10.1101/2021.10.29.466423.

471 Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, Krogh A,

472   Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using

473   deep neural networks. :2022.04.08.487609. DOI: 10.1101/2022.04.08.487609.

474 Ittiprasert W, Knight M. 2012. Reversing the resistance phenotype of the Biomphalaria

475   glabrata snail host Schistosoma mansoni infection by temperature modulation. *PLoS*

476   *Pathog* 8:e1002677. DOI: 10.1371/journal.ppat.1002677.

477 Knight M, Miller AN, Patterson CN, Rowe CG, Michaels G, Carr D, Richards CS, Lewis FA.

478   1999. The identification of markers segregating with resistance to Schistosoma mansoni

479   infection in the snail Biomphalaria glabrata. *Proc Natl Acad Sci U S A* 96:1510–5.

480 Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using

481   repeat graphs. *Nature Biotechnology* 37:540–546. DOI: 10.1038/s41587-019-0072-8.

482 Lu L, Loker ES, Zhang S-M, Buddenborg SK, Bu L. 2020. Genome-wide discovery, and

483   computational and transcriptional characterization of an AIG gene family in the freshwater

484   snail Biomphalaria glabrata, a vector for Schistosoma mansoni. *BMC Genomics* 21:190.

485   DOI: 10.1186/s12864-020-6534-z.

486 Maier T, Wheeler NJ, Namigai EKO, Tycko J, Grewelle RE, Woldeamanuel Y, Klohe K,

487   Perez-Saez J, Sokolow SH, De Leo GA, Yoshino TP, Zamanian M, Reinhard-Rupp J.

488   2019. Gene drives for schistosomiasis transmission control. *PLOS Neglected Tropical*

489   *Diseases* 13:e0007833. DOI: 10.1371/journal.pntd.0007833.

490 Marques DP de A, Rosa FM, Maciel E, Negrão-Corrêa D, Teles HMS, Caldeira RL, Jannotti-

491   Passos LK, Coelho PMZ. 2014. Reduced Susceptibility of a Biomphalaria tenagophila

492   Population to Schistosoma mansoni after Introducing the Resistant Taim/RS Strain of B.

493   tenagophila into Herivelton Martins Stream. *PLOS ONE* 9:e99573. DOI:
494   [10.1371/journal.pone.0099573](https://doi.org/10.1371/journal.pone.0099573).

495 Newton WL. 1955. The establishment of a strain of Australorbis glabratus which combines
496   albinism and high susceptibility to infection with Schistosoma mansoni. *The Journal of*
497   *Parasitology* 41:526–528.

498 Nong W, Yu Y, Aase-Remedios ME, Xie Y, So WL, Li Y, Wong CF, Baril T, Law STS, Lai
499   SY, Haimovitz J, Swale T, Chen S, Kai Z, Sun X, Wu Z, Hayward A, Ferrier DEK, Hui
500   JHL. 2022. Genome of the ramshorn snail Biomphalaria straminea—an obligate
501   intermediate host of schistosomiasis. *GigaScience* 11:giac012. DOI:
502   [10.1093/gigascience/giac012](https://doi.org/10.1093/gigascience/giac012).

503 Paraense WL, Correa LR. 1963. Variation in Susceptibility of Populations of Australorbis
504   glabratus to a Strain of Schistosoma mansoni. *Revista do Instituto de Medicina Tropical de*
505   *Sao Paulo* 5:15–22.

506 Pham TB, Phong CH, Bennett JB, Hwang K, Jasinskiene N, Parker K, Stillinger D, Marshall
507   JM, Carballar-Lejarazú R, James AA. 2019. Experimental population modification of the
508   malaria vector mosquito, Anopheles stephensi. *PLOS Genetics* 15:e1008440. DOI:
509   [10.1371/journal.pgen.1008440](https://doi.org/10.1371/journal.pgen.1008440).

510 Pila EA, Gordy MA, Phillips VK, Kabore AL, Rudko SP, Hanington PC. 2016a. Endogenous
511   growth factor stimulation of hemocyte proliferation induces resistance to Schistosoma
512   mansoni challenge in the snail host. *Proceedings of the National Academy of Sciences*
513   113:5305–5310. DOI: [10.1073/pnas.1521239113](https://doi.org/10.1073/pnas.1521239113).

514 Pila EA, Tarrabain M, Kabore AL, Hanington PC. 2016b. A Novel Toll-Like Receptor (TLR)
515   Influences Compatibility between the Gastropod Biomphalaria glabrata, and the Digenean
516   Trematode Schistosoma mansoni. *PLOS Pathogens* 12:e1005513. DOI:
517   [10.1371/journal.ppat.1005513](https://doi.org/10.1371/journal.ppat.1005513).

518 Richards CS. 1984. Influence of snail age on genetic variations in susceptibility of
519   Biomphalaria glabrata for infection with Schistosoma mansoni. *Malacologia* 25:493–502.

520 Richards CS, Knight M, Lewis FA. 1992. Genetics of Biomphalaria glabrata and its effect on
521   the outcome of Schistosoma mansoni infection. *Parasitol Today* 8:171–4.

522   Richards CS, Shade PC. 1987. The Genetic Variation of Compatibility in Biomphalaria
523   glabrata and Schistosoma mansoni. *The Journal of Parasitology* 73:1146–1151. DOI:
524   10.2307/3282295.

525   Tennessen JA, Bollmann SR, Blouin MS. 2017. A Targeted Capture Linkage Map Anchors the
526   Genome of the Schistosomiasis Vector Snail, Biomphalaria glabrata. *G3-Genes Genomes*
527   *Genetics* 7:2353–2361. DOI: 10.1534/g3.117.041319.

528   Tennessen JA, Bollmann SR, Peremyslova E, Kronmiller BA, Sergi C, Hamali B, Blouin MS.
529   2020. Clusters of polymorphic transmembrane genes control resistance to schistosomes in
530   snail vectors. *eLife* 9:e59395. DOI: 10.7554/eLife.59395.

531   Tennessen JA, Bonner KM, Bollmann SR, Johnstun JA, Yeh J-Y, Marine M, Tavalire HF,
532   Bayne CJ, Blouin MS. 2015a. Genome-Wide Scan and Test of Candidate Genes in the
533   Snail Biomphalaria glabrata Reveal New Locus Influencing Resistance to Schistosoma
534   mansoni. *PLoS Neglected Tropical Diseases* 9:e0004077. DOI:
535   10.1371/journal.pntd.0004077.

536   Tennessen JA, Theron A, Marine M, Yeh JY, Rognon A, Blouin MS. 2015b. Hyperdiverse
537   gene cluster in snail host conveys resistance to human schistosome parasites. *PLoS Genet*
538   11:e1005067. DOI: 10.1371/journal.pgen.1005067.

539   Theron A, Coustau C, Rognon A, Gourbière S, Blouin MS. 2008. Effects of laboratory culture
540   on compatibility between snails and schistosomes. *Parasitology* 135:1179–1188. DOI:
541   10.1017/S0031182008004745.

542   Ulmer MJ. 1970. Laboratory Maintenance of Parasites. In: *Experiments and Techniques in*
543   *Parasitology. MacInnis and M. Voge (eds.)*. W.H. Freeman and Company, 143–144.

544   Waples RS. 2015. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal*
545   *of Heredity* 106:1–19. DOI: 10.1093/jhered/esu062.

546   Webster JP, Gower CM, Blair L. 2004. Do hosts and parasites coevolve? Empirical support
547   from the Schistosoma system. *Am Nat* 164 Suppl 5:S33-51.

548   World Health Organization.Schistosomiasis. *Available at https://www.who.int/news-room/fact-*
549   *sheets/detail/schistosomiasis* (accessed February 1, 2022).

550   Yoshino TP, Wu XJ, Liu H, Gonzalez LA, Deelder AM, Hokke CH. 2012. Glycotope sharing
551   between snail hemolymph and larval schistosomes: larval transformation products alter

552   shared glycan patterns of plasma proteins. *PLoS Negl Trop Dis* 6:e1569. DOI:

553   10.1371/journal.pntd.0001569.

554  Zhang SM, Nian H, Zeng Y, Dejong RJ. 2008. Fibrinogen-bearing protein genes in the snail

555   Biomphalaria glabrata: characterization of two novel genes and expression studies during

556   ontogenesis and trematode infection. *Dev Comp Immunol* 32:1119–30. DOI:

557   10.1016/j.dci.2008.03.001.

558

Figure 1

Backcrossing scheme.

(A) Relative positions of the markers used in marker-assisted backcrossing (actual primer locations are in table S1). Red box = the *PTC2* region, as defined in Tennessen et al. (2020). Marker *0* is in, but near the far upstream end of the *PTC2* region. Distances indicated between markers are approximate. Knight et al.'s (1999) RAPD marker *OPM-04* is approximately 2.5 Mb upstream of *up2*. (B) Illustration of driving the BS90 haplotype into an M-line genetic background. We looked for recombinants between one set of exterior markers (e.g. between *up2* and *up1*). We then backcrossed those individuals and their offspring repeatedly to pure M-line snails until we obtained a crossover between the other pair of exterior markers (e.g. between *dn1* and *dn2*). Then we expanded tanks that were started using the offspring of an MHHHM individual (heterozygous for the three interior markers, but pure M-line at the flanking markers), and then tested for effects of segregation at *PTC2* in the mostly M-line genetic background. We created four of these"B into M" tanks. The same approach was used to produce a single tank segregating for the *PTC2* haplotypes in a mostly BS90 genetic background. We refer to the block of genome between *up2* and *dn2* as the "backcross region", and use "*PTC2* region" to refer just to the ~300 Kb region described in Tennessen et al. (2020) = Red box.

(A)

(B)

Backcross *F1*
to recipient
(M-line)

Additional rounds
of backcrossing
and selection

Expand population
to produce ~1:2:1
on a mostly M-line
genetic background
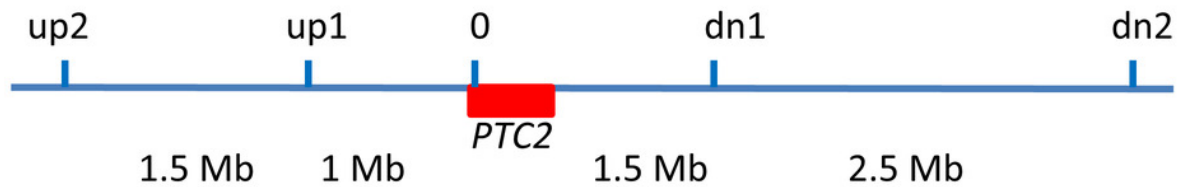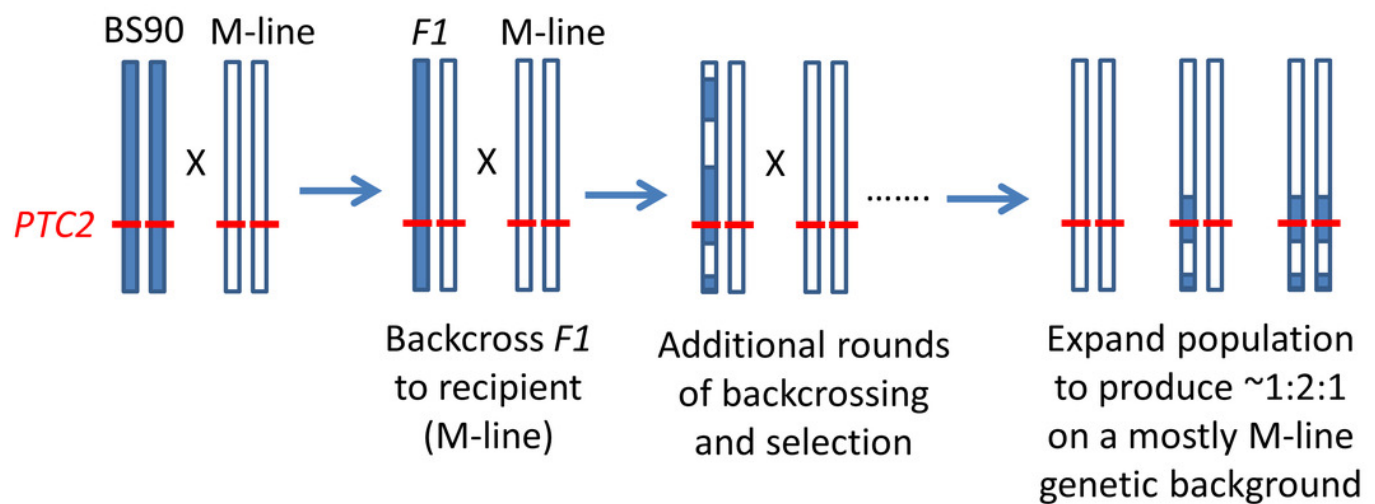
Figure 2

Effect of *PTC2* region (locus *0*) genotypes in four different backcross-into-M genetic backgrounds.

B = BS90 allele. *M* = M-line allele. Number of snails challenged and genotyped for each population are in parentheses. Number of snails infected per number of that genotype are also labelled. Error bars = standard error of the proportion. The population and genotype effects were significant, but the population-by-genotype interaction effects were not. Estimates and P-values for each parameter in the full model are in Table S3. For the genotypic effect, the additive and M-allele-dominant models had similar support and gave odds ratios of 2.2, P = 0.025 and 3.8, P = 0.009, respectively.
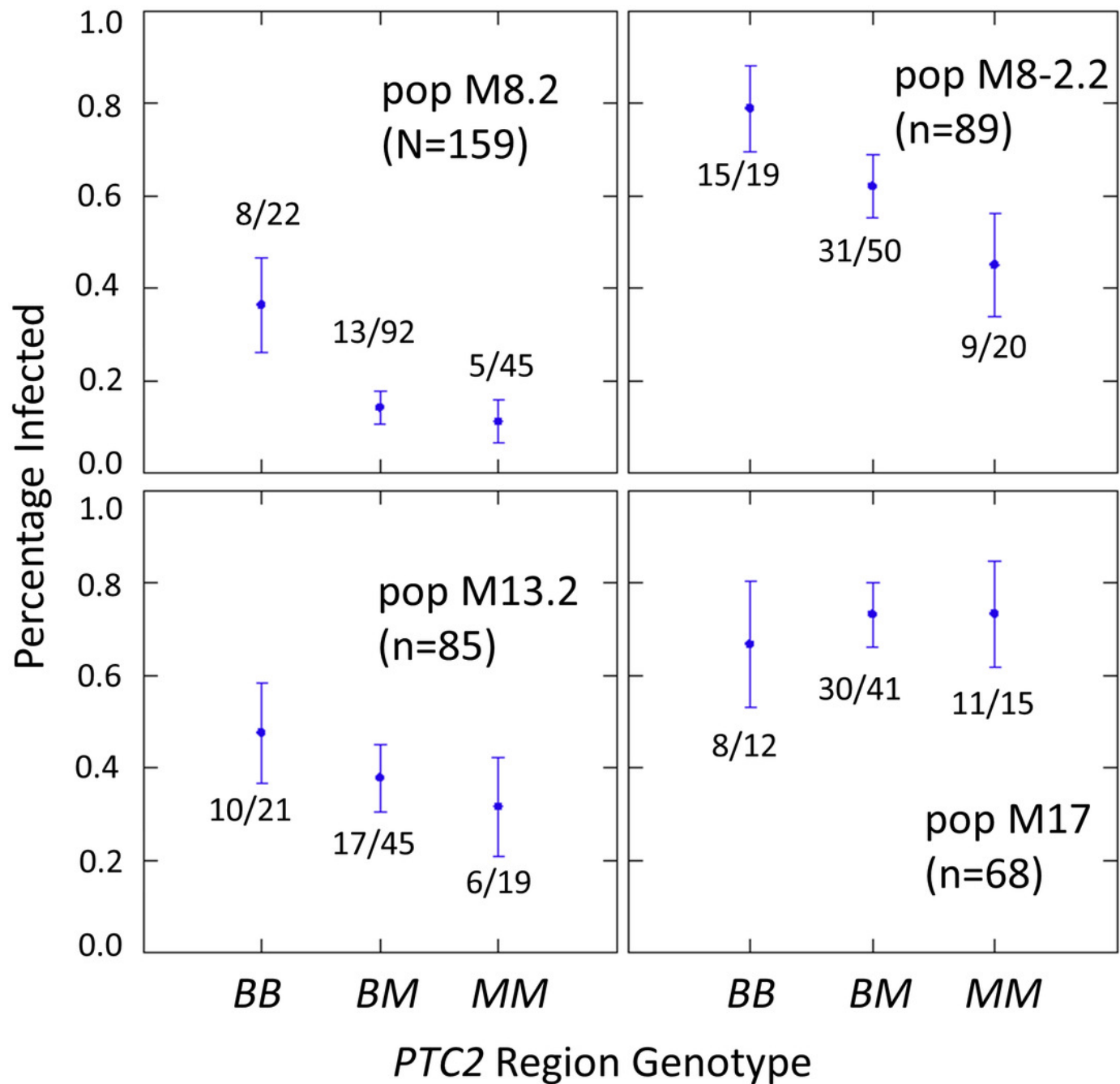
Figure 3

Effects of *PTC2* region (locus *up1*) in the full sample from the F2 population (N = 329).

B = BS90 allele. *M* = M-line allele. Number of snails infected per number of that genotype are labelled. Error bars = standard error of the proportion. The genotype effect was significant, with odds ratios = 1.5, P = 0.041 and 1.8, P = 0.054 for additive and dominant models, respectively. The additive and M-allele-dominant models were equally supported.
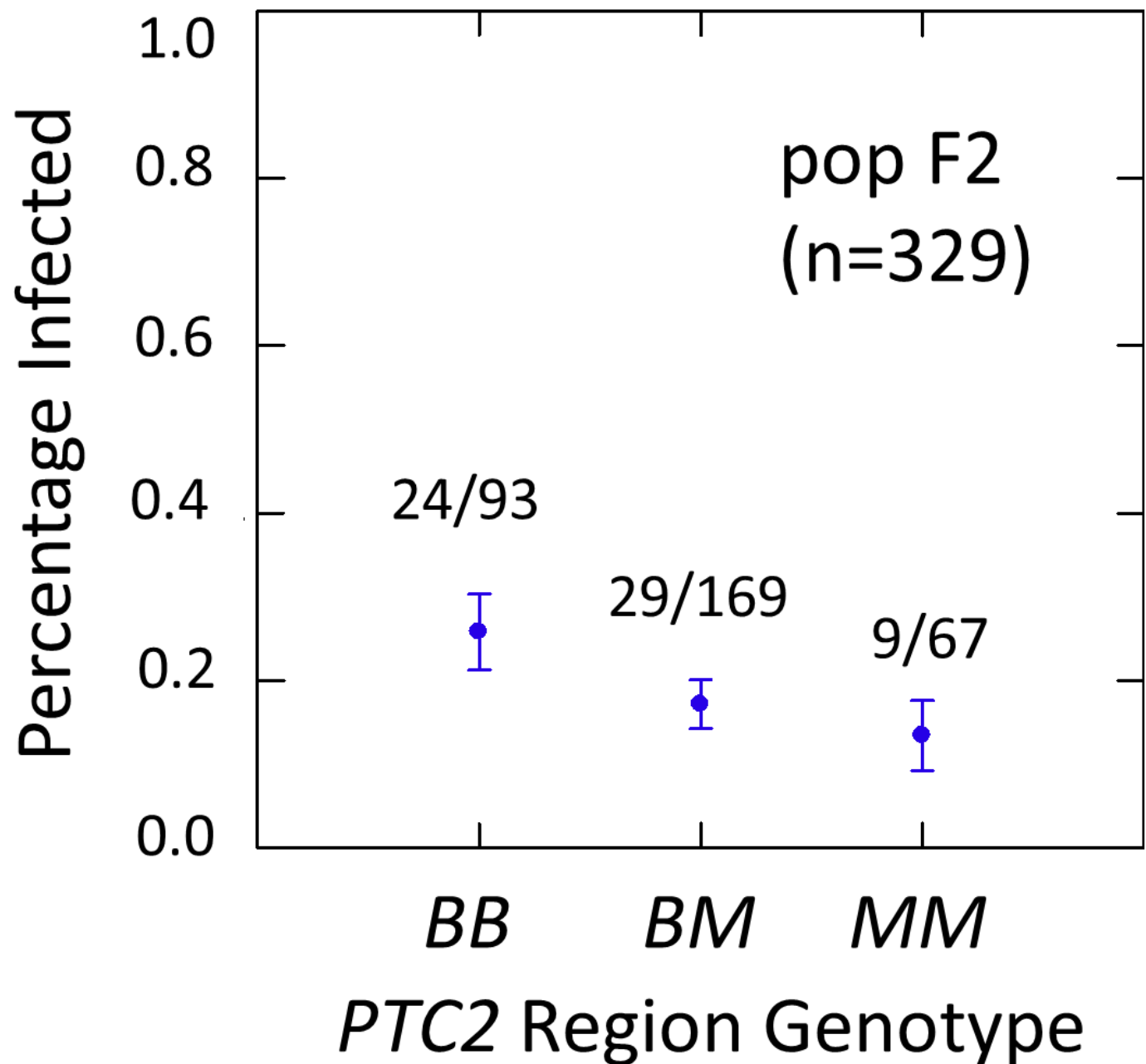
Figure 4

Effects of *PTC2* region (locus *up1*) and locus *OPM-04* in the sub-sample of 55 cases and 55 controls.

Total sample size is slightly less than 110 for each locus owing to PCR failures. *B* = BS90 allele and *M* = M-line allele. Number of snails infected per number of that genotype are labelled. Error bars = standard error of the proportion. The genotypic effect was significant for *PTC2* but not for *OPM-04*. For *PTC2*, the M-allele-dominant model had a better fit (AIC = 147.25 vs. 150.05). The genotypic effect gave odds ratios of 2.0, P = 0.029 and 3.3, P = 0.007 for additive and dominant models, respectively. The average infection rate is higher than in Figure 3 because the 329 individuals in Figure 3 were a random sample from the population.
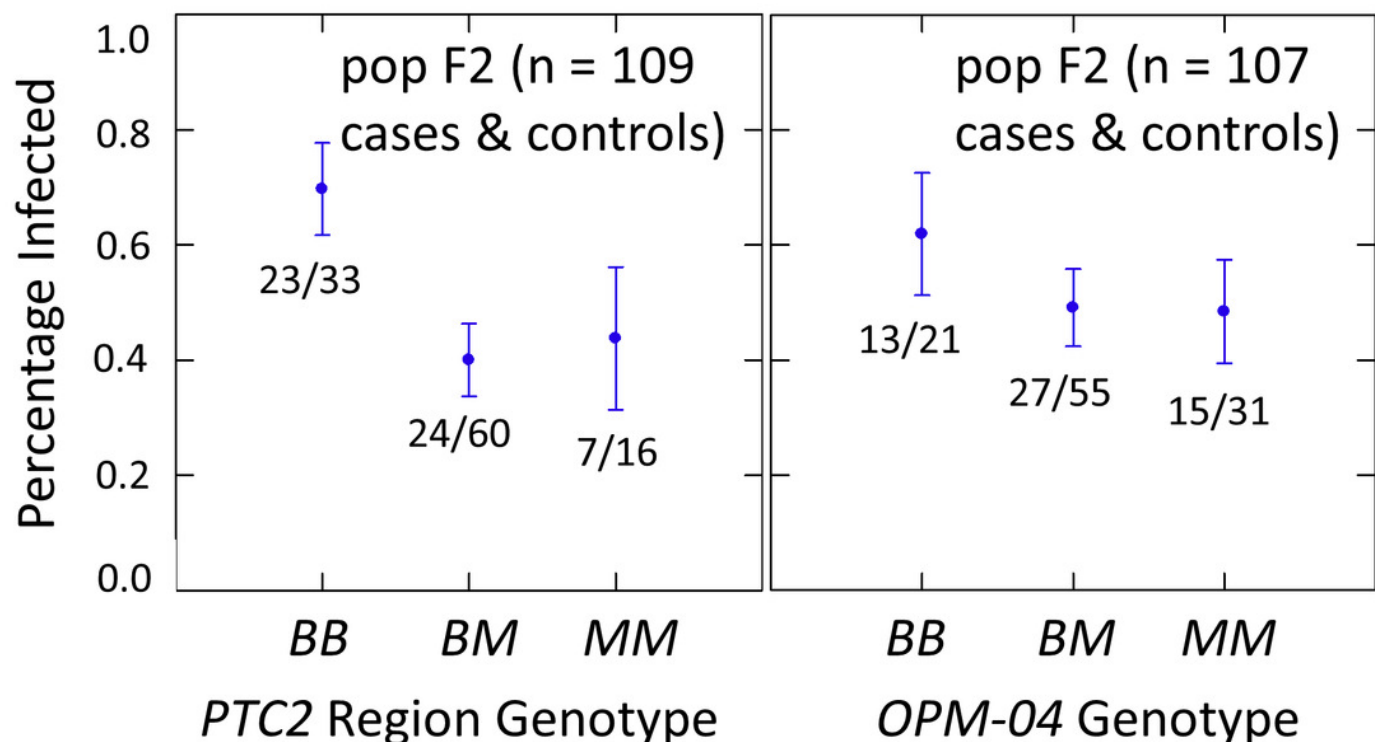
Figure 5

Divergence among *PTC2* coding genes.

Genes are numbered 1-11 as in Tennessen et al. (2020). For M-line, BS90, and the three 1316R haplotypes (*S1*, *S2*, and *R*), squares indicate which genes are present. The *R* haplotype was previously reported to not encode Gene 3 due to lack of observed expression (Tennessen et al. 2020), though an orthologous and syntenic open reading frame does exist, which we indicate here. Color indicates amino-acid similarity to M-line. M-line is nearly identical to *S2*, while BS90 is a mix of *S1*-like and *S2*-like sequence. The region containing genes 6-9 seems the most likely to explain the M-line vs. BS90 difference.
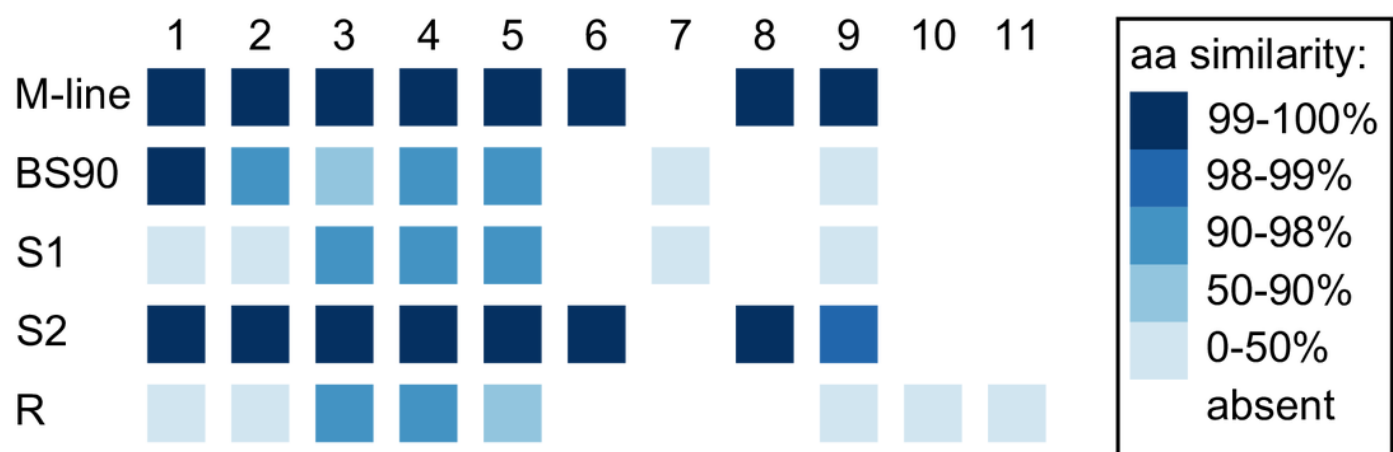
Figure 6

Dot plot for our BS90 and M-line haplotypes in the vicinity of the *PTC2* region.

Peach rectangle = *PTC2* region identified in Tennessen et al. (2020). Each dot represents a 600 bp segment, and color indicates sequence similarity (light blue = 50–90% similarity; medium blue = 90–97.5% similarity; dark blue = 97.5–100% similarity). Mapping markers are indicated (open red circles); *up2* and *dn2* lie beyond the borders of the plot. Candidate genes outside of *PTC2* include a toll-like receptor (TLR) gene and M-line AIG genes (shown with ambiguous x-axis position since they don't occur in BS90). Scale bars of 1 Mb are indicated. In the central ~1Mb region of this plot, BS90 and M-line share almost no sequence similarity except for the first half of the *PTC2*. Gene(s) in the unshared regions are interesting candidates to explain the phenotypic differences observed between haplotypes.