

# UC Irvine

## UC Irvine Previously Published Works

### Title

Compatibility between snails and schistosomes: insights from new genetic resources, comparative genomics, and genetic mapping

### Permalink

<https://escholarship.org/uc/item/86c1f91c>

### Journal

Communications Biology, 5(1)

### ISSN

2399-3642

### Authors

Bu, Lijing  
Zhong, Daibin  
Lu, Lijun  
et al.

### Publication Date

2022-09-01

### DOI

10.1038/s42003-022-03844-5

Peer reviewed

## Compatibility between snails and schistosomes: insights from new genetic resources, comparative genomics, and genetic mapping

Lijing Bu <sup>1,3</sup>, Daibin Zhong<sup>2,3</sup>, Lijun Lu<sup>1</sup>, Eric S. Loker<sup>1</sup>, Guiyun Yan<sup>2</sup> & Si-Ming Zhang<sup>1</sup>✉

The freshwater snail *Biomphalaria glabrata* is an important intermediate host of the parasite *Schistosoma mansoni* that causes human intestinal schistosomiasis. To better understand vector snail biology and help advance innovative snail control strategies, we have developed a new snail model consisting of two homozygous *B. glabrata* lines (iM line and iBS90) with sharply contrasting schistosome-resistance phenotypes. We produced and compared high-quality genome sequences for iM line and iBS90 which were assembled from 255 (N50 = 22.7 Mb) and 346 (N50 = 19.4 Mb) scaffolds, respectively. Using F2 offspring bred from the two lines and the newly generated iM line genome, we constructed 18 linkage groups (representing the 18 haploid chromosomes) covering 96% of the genome and identified three new QTLs (quantitative trait loci), two involved in snail resistance/susceptibility and one relating to body pigmentation. This study provides excellent genomic resources for unveiling complex vector snail biology, reveals genomic difference between resistant and susceptible lines, and offers novel insights into genetic mechanism of the compatibility between snail and schistosome.

<sup>1</sup>Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA. <sup>2</sup>Program in Public Health, College of Health Sciences, University of California, Irvine, CA 92697, USA. <sup>3</sup>These authors contributed equally: Lijing Bu, Daibin Zhong. ✉email: [zhangsm@unm.edu](mailto:zhangsm@unm.edu)

Schistosomiasis, caused by digenetic trematodes of the blood fluke genus *Schistosoma*, is one of world's great neglected tropical diseases, and currently afflicts 237 million people in 78 countries<sup>1</sup>. There is no effective vaccine against schistosomes. Current treatment relies almost exclusively on Praziquantel (PZQ), a drug that has been used for more than 40 years<sup>2</sup>. PZQ-based control programs are not likely to be sufficient to achieve the ultimate goal of disease control and elimination because PZQ-treated patients, especially children, often rapidly reacquire infection<sup>3–5</sup>. Snail control, alone or in combination with other strategies, has been proven to be the most effective means in reducing schistosomiasis transmission in endemic areas<sup>6–8</sup>. Strategies currently in use for snail control include biocontrol using competitors or predators, modification of snail habitats, and application of molluscicides. Each approach, however, has been found to have significant limitations, and is not sustainable. Genetics-based vector snail control has long been advocated as a novel biocontrol strategy of schistosomiasis<sup>9,10</sup> and such a strategy has already shown great promise in control of insect vector-transmitted diseases<sup>11,12</sup>, but not in schistosome vector snails. Studies of the vector snail *Biomphalaria tenagophila* in Brazil<sup>13,14</sup> provide supportive field evidence for the validity of the approach, though the underlying mechanism for resistance remains to be explored in *B. tenagophila*. To accomplish the exciting goal of schistosomiasis control, a critical understanding of vector snail biology, particularly genetic mechanisms underlying snail resistance to schistosomes, is required<sup>15,16</sup>.

*Biomphalaria glabrata* – *Schistosoma mansoni* has been applied for studies of snail-schistosome interactions for more than half a century<sup>17–20</sup>. The Neotropical species *B. glabrata* is an important intermediate host of the parasite *S. mansoni*, a causative agent of human intestinal schistosomiasis. Two *B. glabrata* strains, M line and BS90, have been most used for research and supplied by NIH-funded Schistosomiasis Resource Center of Biomedical Research Institute (BRI) for world-wide research ([www.afbr-bri.org/schistosomiasis](http://www.afbr-bri.org/schistosomiasis))<sup>21</sup>. In general, albino M line is susceptible to *S. mansoni*, whereas pigmented BS90 is resistant to the parasites. M line was selected from early crosses between albino Brazilian strain and pigmented Puerto Rican snails in 1950s<sup>17</sup>. The pigmented BS90, also referred to frequently as the Salvador strain, was isolated from field populations in 1960s in Salvador, Brazil, which were found to be resistant to infection with both local and allopatric *S. mansoni*<sup>22</sup>. By using the two strains, most of the past work has focused on immunological responses of *B. glabrata* to digenetic trematodes, particularly *S. mansoni*<sup>18–20</sup>. Only a few genetics studies concerned with genetic mapping of schistosome resistance in *B. glabrata* have been reported. Knight et al.<sup>23</sup> identified two schistosome-resistant loci (OPM-04 and OPZ-11) using adult F2 offspring produced from crosses between non-inbred M line and BS90 *B. glabrata* snails. More recently, three new resistant loci, GRC, RADres, and PTC2, have been uncovered from juvenile snails of Guadeloupe and 13-16 R1 strains of *B. glabrata*<sup>24–26</sup>. Although these efforts have significantly improved our knowledge of snail resistance in general, fundamental genetic mechanisms underlying snail resistance to schistosomes remain elusive. For example, all the above-mentioned resistance loci have been found in different genomic regions. Clearly, further investigations using advanced genetic and genomic resources and more powerful genotyping technologies are needed.

We have developed a new snail model consisting of two homozygous lines with sharply contrasting phenotypes of schistosome-resistance, namely iM line and iBS90, which were derived from the M line and BS90 snails as described. To empower this genetic resource, we have sequenced, assembled, and annotated the genomes of the two lines. The quality of the two genomes

generated from the two lines representing two distinct phenotypes (resistance and susceptibility) has been significantly improved compared to the first pioneering reference genome of *B. glabrata* reported<sup>27</sup>. To exploit these unique genetic and genomic resources for better understanding genetic mechanisms of snail resistance to schistosomes, we produced F2 segregating population using the two lines as parents, constructed a high-quality linkage map, and identified three novel quantitative trait loci (QTLs) involved in snail resistance and pigmentation using double digest restriction-site associated DNA sequencing (ddRADseq)<sup>28</sup>. Taken together, our findings (two homozygous lines, two genome sequences, one linkage map, and three QTLs) presented in this paper provide valuable genetic and genomic resources for better understanding complex biology of vector snails and offers insights into genetic basis underpinning snail resistance or susceptibility to schistosomes.

## Results

**Generation of two new *B. glabrata* lines.** Here we report a new snail model that consists of two new lines, namely iM line and iBS90, which were selected from M line and BS90, respectively (for details see “Methods”; Supplementary Fig. 1a). The two lines possess two important properties, homozygosity of their genomes and sharply contrasting schistosome-resistance phenotypes. K-mer analysis<sup>29</sup> suggests the iM line (81 consecutive generations of selfing) and iBS90 (39 consecutive generations of selfing) are homozygous (Supplementary Fig. 1b). By the criterion of producing cercariae after exposure to *Schistosoma mansoni* miracidia or not, it has been repeatedly confirmed that iM line is fully susceptible to *S. mansoni* (produce cercariae) whereas iBS90 is completely resistant to the parasite (no cercariae produced). Snails of the iM line are albinos whereas iBS90 snails have pigmented bodies and eyes.

**Sequencing, assembly, and annotation of the two lines.** For each homozygous line, two individual snails were used for whole genome sequencing, one for Illumina, and the other for PacBio sequencing (Table 1). The assembled genome size is 871 for iM line (255 scaffolds; N50 = 22.7 Mb) and 885 Mb for iBS90 (346 scaffolds; N50 = 19.4 Mb) (Table 1). Benchmarking universal single-copy orthologue (BUSCO) analysis shows that the calculated completeness of the assembled genome for both iM line and iBS90 is 96% based on a Metazoa dataset of 954 core genes<sup>30</sup>. The overall statistical data indicate that the quality of the iM line genome assembly is better than that of iBS90, so we used the iM line genome as reference for our linkage group analysis and genetic mapping presented in this paper. Table 1 provides detailed information of the two new genome assemblies as well as the first reference genome for *B. glabrata* of the BB02 laboratory strain derived from a Brazilian field isolate<sup>27</sup>.

A total of 35,015 gene models for iM line were predicted using annotated BB02 proteins and RNA-Seq data available in public databases. Among four functional annotation analyses applied, 28,554 genes (81.54% of the 35,015) were annotated from at least one of the four sources: (1) BLASTp against high-quality manually curated Uniports database (15,263 counts for 44% of the total); (2) BLASTn against NCBI non-redundant nucleotide (NT) database (23,781 count for 68% of the total); (3) NCBI non-redundant protein (NR) database (23,673 count for 68% of the total); and (4) domain prediction from InterProScan and IgSF domain models (23,949 count for 68% of the total) (Supplementary Fig. 2a). Using the same approach, a total of 38,516 gene models were predicated in iBS90 (Supplementary Fig. 2b). We found 21,368 ortholog groups between 23,142 iM line genes and 23,318 iBS90 genes (not all genes have ortholog

**Table 1 Summary statistics of the genome assembly of iM line and iBS90 in comparison with reference genome of BB02 strain.**

	iM line	iBS90	BB02
<i>Genome sequencing and raw data</i>			
Sequencing platform	Illumina	Illumina	Sanger+454+Illumina
Number of reads	410,797,214	439,992,466	
Mean of read (bp)	150 × 2	150 × 2	
Median of read (bp)	150 × 2	150 × 2	
Total of bases (Mbp)	123,239	131,998	
Coverage	149	154	79
<i>Sequencing platform</i>			
Sequencing platform	PacBio	PacBio	
Number of reads	7,308,853	6,198,071	
Mean of read (bp)	13,712	12,189	
Median of read (bp)	24,502	25,197	
Total of bases (Mbp)	100,217	75,550	
Coverage	120	87	
<i>Assembly statistics</i>			
Total length (bp)	870,959,050	885,111,083	916,377,450
Number of scaffolds	255	346	331,400
Mean scaffold length (bp)	3,415,526	2,558,125	2,765
Longest scaffold length (bp)	60,252,455	35,282,478	2,183,814
Shortest scaffold length (bp)	1,308	1,344	200
Scaffold N50	22,698,051	19,395,504	48,059
Scaffold L50	13	18	3,093
GC content	36.14%	36.10%	36.19%
BUSCO	96%	96%	88%
Total length (>=10 Kbp)	870,886,136	885,159,316	644,167,852
Number of contigs (>=10 Kbp)	238	336	11,103

Note: Complete and fragmented BUSCOs found from OrthoDB v10 Metazoa datasets ( $N = 954$ , [https://busco.ezlab.org/list\\_of\\_lineages.html](https://busco.ezlab.org/list_of_lineages.html)). Data of BB02 strain were from Adema et al.<sup>27</sup>.

between the two lines). In addition, the iM line and iBS90 genomes contain 42.95% and 55.17% repetitive elements, respectively; more unclassified repeats were also found in iBS90 genome (Supplementary Fig. 3, Supplementary Table 1).

**Structural variation of the genomes between the two lines.** A large number of small structural variations (SV) was found between the two genomes (Fig. 1). The total size of genomic fragments involved in SV was 50 Mb, with a size ranging from 536 bp to 2.8 Mb (median of 4.6 Kb), which covers 5.8% of the whole genome (Supplementary Table 2, Fig. 1). The first quartile (25th percentile) is 1.5 Kb and the third quartile (75th percentile) is 4 Kb, suggesting overall SVs are small. Large-scale sub-chromosomal level rearrangements were not detected between the two genomes. Detailed breakdown of sizes for each SV type is presented as a violin plot (Fig. 2). The SVs were found in 60 scaffolds, overlapping the exons of 1446 protein-coding genes (Supplementary Data 1). This suggests that differences with respect to the schistosoma-resistance trait between the two lines may not result from a large-scale structural variation.

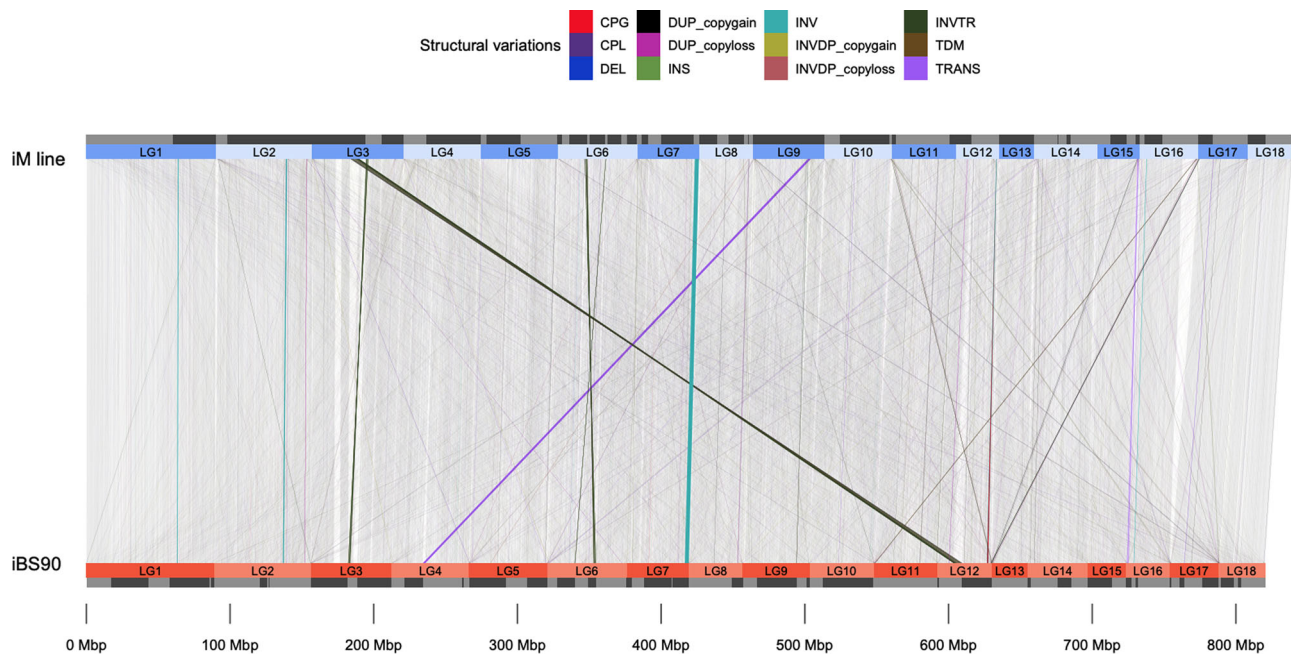
**Phenotypic analysis of F2 progenies bred from the two lines.** By using the iM line (albino) and iBS90 line (pigmented) as starting parents, we generated a F2 segregating population for genetic studies (Supplementary Fig. 1a). Of 869 F2 juvenile snails tested for schistosoma-resistance phenotype, 421 were found to be resistant to *S. mansoni* and 448 were susceptible. A Chi-square goodness of fit demonstrated that the ratio of resistance to susceptibility (or vice versa) did not display as ratio of 3:1 ( $P < 0.05$ ). This suggests schistosoma resistance in juvenile snails is not likely to be controlled by a single gene (Chi-square value for the homogeneity:  $\chi^2 = 18.78$ ,  $P < 0.01$ ), which is in an agreement with previous studies using classic genetic crossings<sup>31,32</sup>. In addition, body color (pigmented or albino) or eye color (pigmented snails also have black eyes) of the same F2 snails was also recorded, and

674 pigmented and 195 albino F2 progenies were obtained ( $\chi^2_{3:1} = 3.04$ ,  $P = 0.08$ ), indicating a single gene or locus is involved in controlling snail pigmentation (Chi-square value for homogeneity:  $\chi^2 = 28.30$ ,  $P < 0.0001$ ). There was no evident association between albinism/pigmentation and resistance or susceptibility status to *S. mansoni*.

**Construction of a linkage map.** From 126 ddRAD (*EcoR* I and *Msp* I) libraries generated, 122 F2 snails and 4 parent snails (2 from either iM line or iBS90), were successfully sequenced using the Illumina HiSeq platform (Supplementary Data 2). The cleaned reads were mapped to iM line and iBS90 genomes, allowing the identification of 90,898 and 92,143 high-quality single nucleotide polymorphisms (SNPs), respectively (Supplementary Fig. 4a, b, c, d). The transitions (C/T and A/G) over transversions (A/C, G/T, A/T, and C/G) ratio was 1.86 in SNPs based on the two genomes. A total of 1482 SNPs (1.63%) for iM line and 1546 (1.68%) for iBS90 were exonic, while the remaining SNPs were found in intronic and intergenic regions. High proportions of shared detection and comparable SNP markers were revealed upon independent assessment of each genome, suggestive of high-quality genome assembly of the two lines. To simplify the study, subsequent analyses of SNPs were performed based on the iM line genome.

Of the 122 F2 snails, 116 were retained after removing individuals with incomplete or low-quality genotype information. Of 90,898 biallelic SNPs, 17,392 were retained after filtering or removing SNPs with high missing data (>35%), apparent segregation distortions ( $P < 0.001$ ) and genotype error, and loci with high levels of pairwise linkage disequilibrium (LD). After tagging of 17,392 SNPs using HaploView with 3.0 of LOD (logarithm of the odds) and  $r$ -squared correlation coefficients (<0.8), and 100 bp of minimum distance between loci, a set of 1146 reliable SNPs was selected for linkage map construction. Of 1146 segregating SNP loci, 996 (~87%) were assembled into the





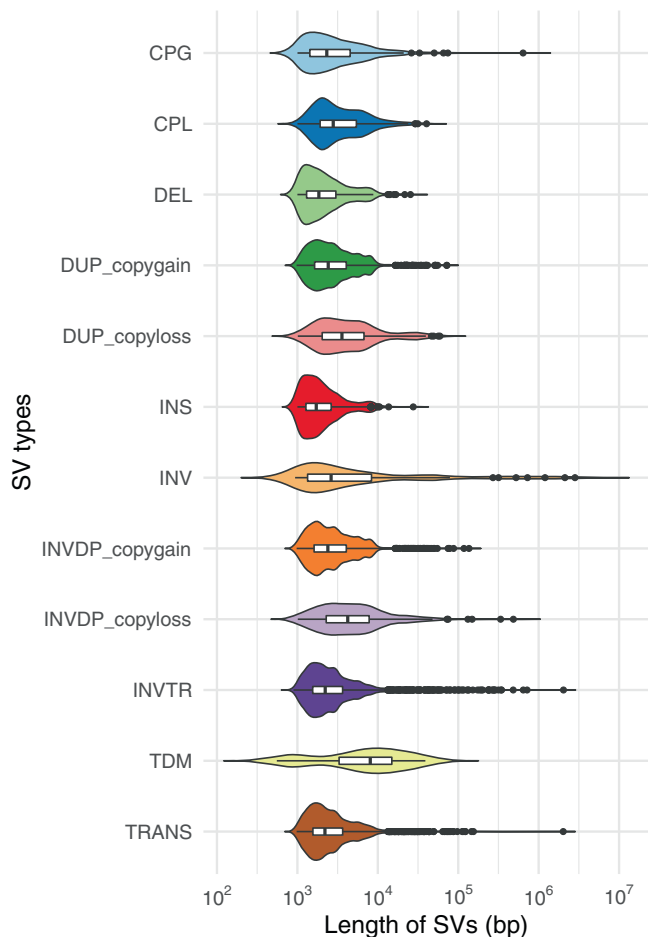
**Fig. 1 Synteny plot and structural variations (SVs) between assemblies of iM line and iBS90.** The numbers of linkage group were assigned based on the sum of lengths of the scaffold on each linkage group. The linkage groups are marked in colored boxes for iM line (blue on top) and iBS90 (orange at bottom). Separated scaffolds inside each linkage group were marked out as gray and black boxes. Synteny blocks between two genomes are laid out in gray blocks in the background. SV types were highlighted in different colors. Cpg: copy gain in query, Cpl: copy loss in query, DEL: deletion in query, DUP: duplicated region, HDR: highly diverged regions, INS: insertion in query, INV: inverted region, INVDP: inverted duplicated region, INVTR: inverted translocated region, NOTAL: un-aligned region, SNP: single nucleotide polymorphism, SYN: syntenic region, TDM: tandem repeat, TRANS: translocated region.

18 large linkage groups (LGs) with the most numbers of markers (>32 SNP loci) at a LOD score of 9.0 and a maximum recombination frequency of 0.14 (Supplementary Data 3). Among the remained 150 SNP loci, half of them ( $n = 75$ ) are either orphan markers unlinked to other makers or located in linkage groups with markers less than 10, so these 150 SNP loci were excluded from linkage map analysis. The numbers for linkage groups were assigned based on the total length of scaffolds on the linkage group. The markers on the genetic linkage map presented the same sequential order as that presented in the physical map from the iM line reference genome. The final genetic linkage map consisted of 18 LGs, which contain 996 SNP markers and span 1664.7 Kosambi cM with an average distance between markers across the 18 linkage groups of 1.73 cM (Fig. 3). The genetic distance of linkage groups ranged from 49.1 cM (LG15) to 164.7 cM (LG1). The number of SNP markers mapped to each linkage group varied from 32 markers on LG14 to 93 markers on LG1, with an average of 54 SNPs per linkage group. The 18 LGs that represent the 18 haploid chromosomes characteristic of *B. glabrata* ( $2n = 36$ )<sup>33</sup>, were composed of 64 scaffolds, which cover 96% of the assembled iM line genome (Fig. 3, Supplementary Data 3, 4).

Given an estimated size of 871 Mb in the genome assembly of iM line (Table 1), the current linkage map is  $\sim 0.52$  Mb/cM. The recombination blocks were clearly defined in the graphical representation of genotypes (Fig. 4), indicating that the genotypic dataset used for construction of the final linkage map was well-corrected and suitable for linkage mapping<sup>34</sup>. In the pairwise recombination fraction plot, some markers had high recombination frequencies (>0.5), but low LOD scores (Supplementary Fig. 5a). The heatmap of estimated recombination and LOD scores of mapped markers showed a low level of off-diagonal association between marker pairs and consistency of square-like heat blocks across the diagonal line within chromosomes,

suggesting this is a good-quality linkage map (Supplementary Fig. 5b)<sup>35</sup>. Of the 996 mapped SNP markers, 121 (12.1%) showed significant ( $P < 0.01$ ) distortion from the 1:2:1 segregation ratio in the F2 population, with heterozygote deficiency (LG11 and LG13) or heterozygote excess (LG1 and LG8) (Fig. 5).

**QTLs conferring schistosome-resistance and body pigmentation.** Using a genome-wide LOD significant threshold of 4.0 and permutation test<sup>36</sup>, two QTLs (qRS-2.1 and qRS-5.1) associated with schistosome resistance/susceptibility were detected on two LGs or chromosomes (LG2 and LG5) using simple interval mapping (SIM) (Fig. 6, Table 2). Relatively high LOD scores were noted in LG4 and LG10, but not selected because they did not meet the criterion (i.e.,  $\text{LOD} > 4$ ). The peak of largest effect (qRS-5.1) was located between markers m8-12003253 and m8-13368092, with a genetic distance of 0.50 cM or physical distance of  $\sim 1.36$  Mb. This QTL had an additive effect of 0.30 for increased resistance to parasite infections when an individual harbored the allele from a resistant parent. The second QTL (qRS-2.1) is located between the markers m2-36225505 and m2-37212557 and has a genetic distance of 1.43 cM or physical distance of  $\sim 0.99$  Mb. This QTL had a dominance effect of 0.41 for reduced resistance to parasite infections when an individual harbored the allele from a susceptible parent. In addition to using lack of cercarial shedding as a resistance phenotype as just described, we also examined the abundance of parasite DNAs/reads present in the snail body as an indication of schistosome resistance. For a given snail, the more schistosome reads counted in its DNA sample, the more susceptible to schistosomes the snail was considered to be. Therefore, reads that were unmapped to the snail genome but mapped to the *S. mansoni* genome<sup>37</sup> were used as an indication for abundance of parasites present in a snail (Supplementary Fig. 6, Supplementary Data 2). Remarkably, two QTLs (qsm-2.1 and qsm-5.1) revealed based on the number of



**Fig. 2 Violin plots of SVs lengths distributions between iM line and iBS90 genome assemblies.** CPG: copy gain in query, CPL: copy loss in query, DEL: deletion in query, DUP: duplicated region, INS: insertion in query, INV: Inverted region, INVDP: inverted duplicated region, INVTR: inverted translocated region, SYN: syntenic region, TDM: tandem repeat, TRANS: translocated region.

schistosome reads are consistent with the patterns observed by the resistance phenotype (cercariae shedding) for both QTL regions (qRS-2.1 and qRS-5.1), except for varying confidence intervals, supportive of the high reliability of the two QTLs identified (Fig. 6; Table 2). The size of the two QTLs (qRS-2.1 and qRS-5.1) covers an interval of 0.5–1.43 cM in genetic distance or 0.99–1.36 Mb in physical distance, while the 95% confidence intervals of QTLs span large chromosomal regions ranging from 2 to 8 cM genetic distance or 4.03–21.36 Mb physical distance. Interestingly, both QTLs locate in centromeric and pericentromeric regions inferred from the genetic map as corresponding to large, low or zero recombination regions<sup>38</sup>, suggesting centromeric/pericentromeric regions may play a role in schistosome resistance/susceptibility in *B. glabrata*.

Regarding body pigmentation, one major QTL (qBC-13.1) was detected. This QTL was found on LG13 and has very high LOD score (8.92) with an additive effect value of 0.34 (Fig. 6; Table 2). No minor QTLs (PVE < 10%) was detected for snail pigmentation.

To validate the association between the significant SNPs and snail phenotypes, 94 F2 snails (randomly selected from those used for ddRADseq) and 48 F2 snails (not used for ddRADseq) were analyzed by Sanger bi-directional sequencing. Eleven markers from the three loci (qRS-2.1, qRS-5.1, and qBC-13.1)

and one previously reported (i.e., OPM-04, GenBank accession: AF078109<sup>23</sup>), were successfully genotyped by Sanger sequencing. The odds ratio (OR, a measure of strength of association) of the 12 markers ranged from 2.27 to 11.12 ( $P < 0.01$ ) for technical validation and 3.39 to 8.10 ( $P < 0.05$ ) for biological validation (Supplementary Tables 3, 4). The independent analysis validates these QTL loci identified.

**Genes presented in the QTLs.** Based on the genome annotation of iM line presented in this study, 1,348 genes were found in the two loci within the 95% confidence intervals for the QTL regions (a region of ~7.95 Mb on LG2 and ~21.36 Mb on LG5). Among the 1348 genes, a total of 328 were observed to be associated with at least one significantly differential SNP (a total 1315 SNPs) (Supplementary Data 5, 6). Of these 328 genes, 97, 222, and 9 genes were from qRS-2.1, qRS-5.1, and qBC-13.1, respectively. For the 97 genes from qRS-2.1 (448 SNPs), six genes (e.g., epithelial membrane protein and vacuolar protein 8) that possess non-synonymous coding mutations (Supplementary Table 5) and several genes (e.g., neuronal acetylcholine receptor and phosphoribosyl pyrophosphate synthase-associated protein) were found with mutations within putative regulatory regions (intron, upstream and downstream protein-coding sequences) (Supplementary Table 6). For the 222 genes (804 SNPs) located in qRS-5.1, eight had non-synonymous coding changes (e.g., receptor-type tyrosine-protein and serine/threonine-protein kinase, Supplementary Table 5) and several genes (e.g., ferric-chelate reductase, peroxidase, and orexin receptor) had mutations within the regulatory regions (Supplementary Table 6). Gene ontology (GO) analyses for both resistance QTLs showed that those genes were mainly enriched in specific processes, including the highest enrichments of “binding” term for molecular function and “cellular process” term for biological process. In the cellular component, the highest enrichment was found in membrane proteins (Supplementary Fig. 7a, b).

With respect to the pigmentation locus qBC-13.1, all 9 genes associated with a total of 63 SNPs, including a neuropeptides capsa receptor, displayed a single Gln239Lys non-synonymous mutation and several genes (e.g., ATP-dependent DNA/RNA helicase DHX36 and ADAM family mig-17) were found to possess mutations in regulatory regions, but not coding regions. A large deletion was not observed in the coding regions of the genes. The highly enriched genes in pigmentation locus are involved in catalytic activity and cellular processes (Supplementary Fig. 7c).

## Discussion

The draft *B. glabrata* genome was reported in 2017<sup>27</sup> and provides an invaluable resource for gene discovery and data mining. This pioneering reference genome sequence was generated from many snails using multiple technologies available at the time, including Sanger-, 454-, and Illumina-sequencing. As such, the sequence assembly was highly fragmented (N50 = 48.1 Kb; number of scaffolds = 331.4 K) (Table 1). Recently, three significantly improved and assembled, but not annotated, genome sequences of the *B. glabrata* 13-16-R1 strain that were used for genetic mapping<sup>26</sup> were released to GenBank (Assembly number: ASM1452496, ASM1452502, and ASM1452495). The numbers of scaffolds for the three genomes are 927, 2718, and 3492, respectively, and corresponding sizes of assembled genomes are 852, 811, and 768 Mb. The two genomes we report here were built on 255 (iM line) and 346 (iBS90) scaffolds, each derived from two homozygous individuals. Using homozygous lines circumvents drawbacks associated with the use of heterozygous strains, which impair production of high-quality genomes<sup>39</sup>. Although extreme inbreeding may result in homozygous

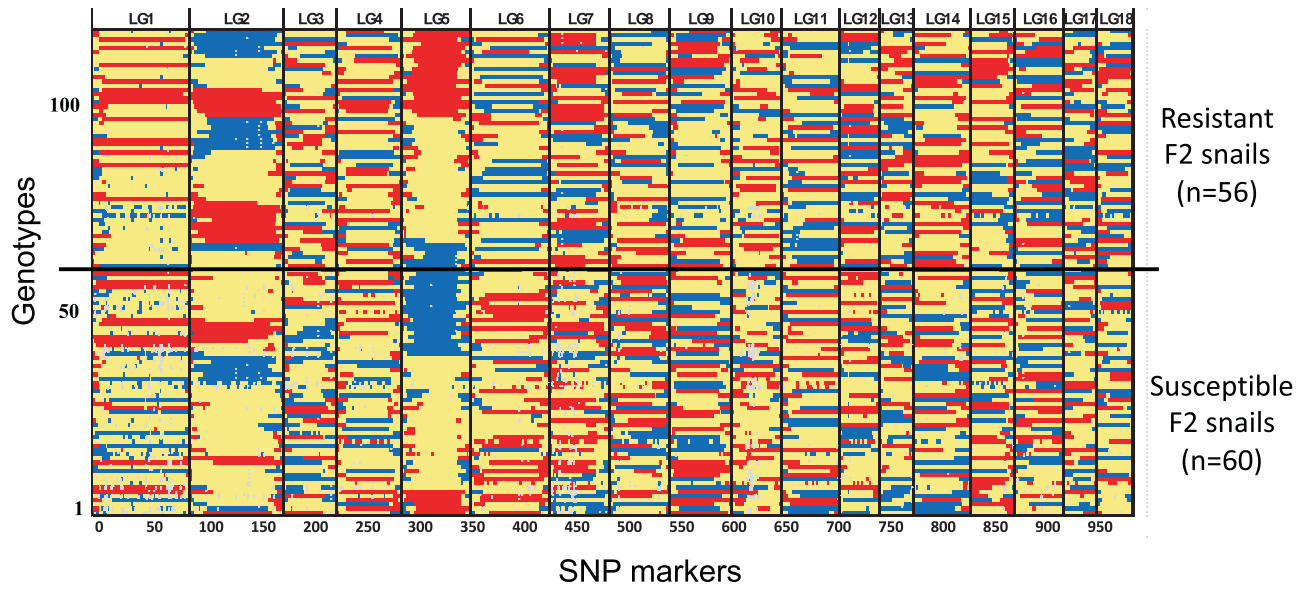


**Fig. 3 A genetic linkage map based on a F2 offspring bred from iM line and iBS90.** The markers were labeled by the scaffold # followed by the physical position of the SNP. Markers in red represent those markers used for validation by Sanger sequencing.

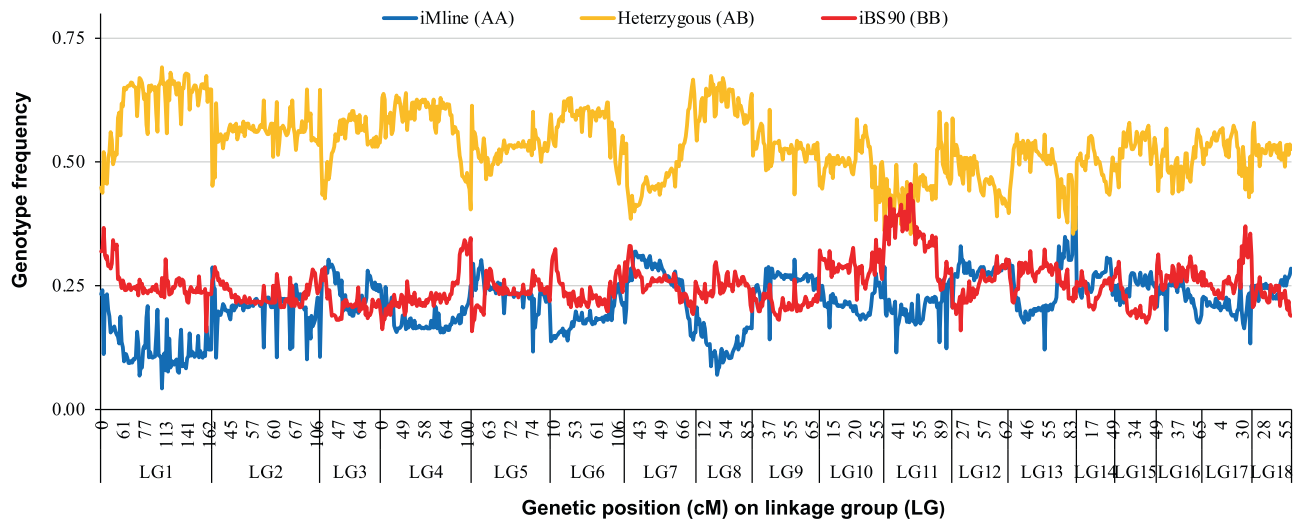
representation of deleterious alleles, which may lead to undesirable traits and affect fitness, our homozygous lines maintain contrasting phenotypes with respect to schistosome resistance, preserving the essential trait for our studies of snail-schistosome interactions. Therefore, the two new lines with sharply contrasting vector competence phenotypes and high-quality genome sequences provide valuable genetic and genomic resources for studies of vector snail biology and host-parasite interactions. The unique genetic resources (iM line and iBS90) along with their genome sequences and the data used for linkage and QTL mapping will be provided or released to the scientific community as soon as this work is published. As the quality of the susceptible iM line genome is better than that of the resistant iBS90 genome or of other *B. glabrata* genomes (Table 1<sup>26,27</sup>), the iM line genome can serve as an improved reference genome for *B. glabrata*. An additional step we discuss below to make the iM

line even more valuable as a reference genome for schistosome-miasis studies, is generation of chromosome-level assemblies. ddRADseq, a powerful tool for large-scale, high-throughput genotyping, has been used effectively to develop linkage maps and for genetic mapping in many non-model genetic organisms<sup>40–42</sup>. We employed ddRADseq to identify genome-wide SNP markers and then constructed a highly saturated genetic linkage map using an F2 segregating population derived from two homozygous snail parents. The 18 linkage groups, representing the 18 haploid chromosomes ( $2n = 36$  in *B. glabrata*) were, in general, consistent with the 18 large LGs reported previously<sup>43</sup> although a few scaffolds from a given LG are not exactly matched (Table 3, Supplementary Table 7). Different from the previous version consisting of 24 LGs and 56% genome coverage<sup>43</sup>, our version built on the iM genome contained 18 LGs and covers 96% of the *B. glabrata* genome.





**Fig. 4 Graphical representation of genotypes on 18 linkage groups.** Blue, red, and yellow colors represent alleles that are derived from resistant iBS90, susceptible iM line and heterozygous (from both parents), respectively. The white color indicates missing information. Black vertical lines indicate the boundaries between LGs. The y-axis was sorted by genotypes of QTL regions on LG5 and LG2 to show the proportion of genotypes in resistant group (upper) and susceptible group (lower).



**Fig. 5 The proportion of contributing alleles from iM line (AA) and iBS90 (BB) for each marker on the 18 linkage groups.** Parental allele contributed by iM line and iBS90 is represented in blue and red, respectively, whereas heterozygous contributed by both parents is indicated in yellow.

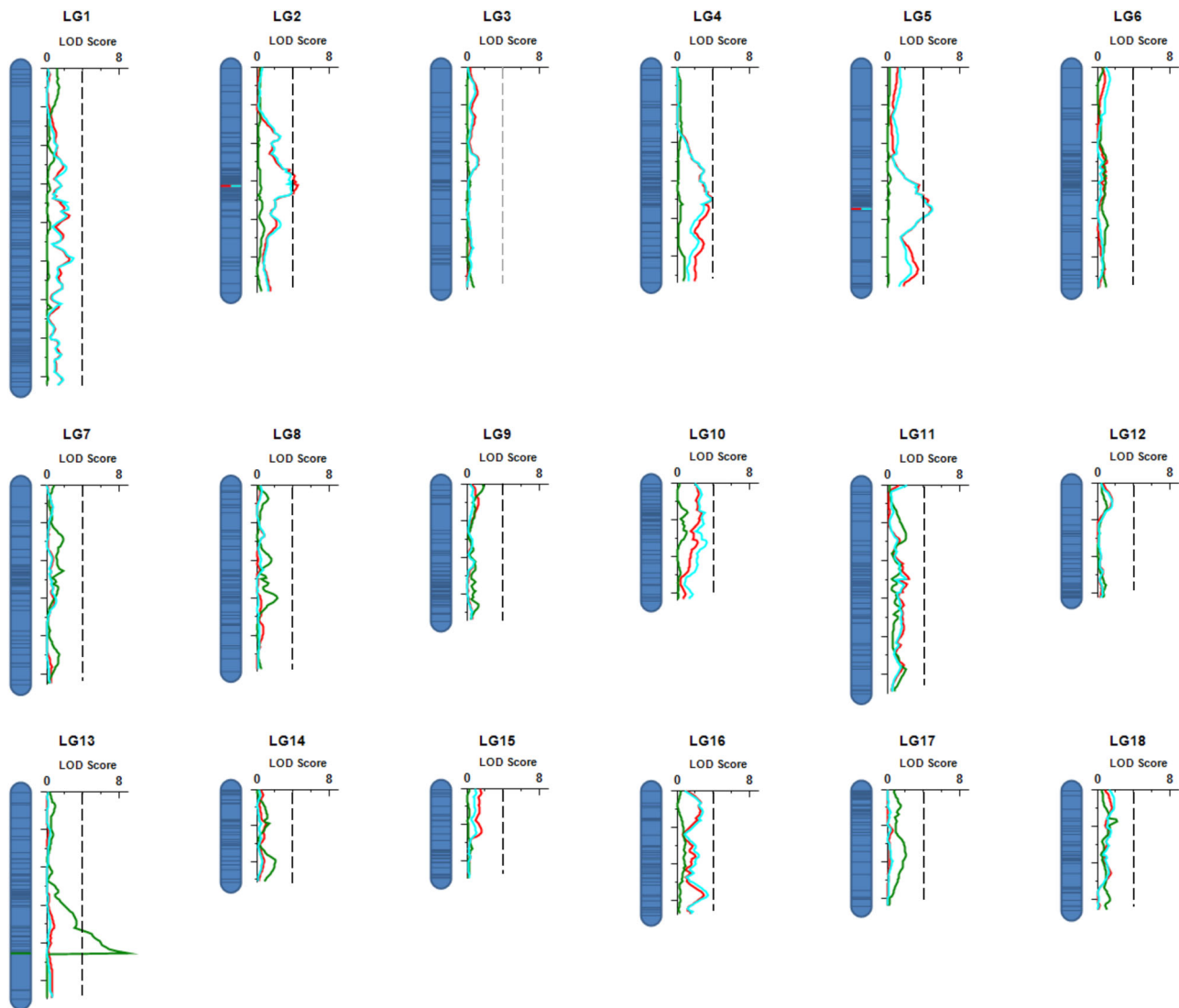
We observed a small portion (12.1%) of SNP markers that did not follow the Mendelian segregation ratio (1:2:1) in the F2 population. Most of those markers located on LG1 and LG8, showed a significant excess of heterozygosity. The exact reason leading to heterozygote excess is not known, but this phenomenon is commonly found in animal and plant species and believed to act as an important evolutionary force<sup>44</sup>. The distorted markers were generally clustered and displayed the same skew direction, suggesting segregation distortion loci under selection on linked marker loci<sup>45</sup>. Consequently, zygotic selection, such as the differentiation of zygote viability or heterosis, may be the genetic mechanism for the observed segregation distortion<sup>46</sup>. In addition, markers on LG11 in the iBS90 genome were skewed towards homozygotes, implying that iBS90 homozygotes were preferred under selection.

Revealing genetic mechanisms of schistosome-resistance in vector snails is a most important motivation for studies of snail-

schistosome interactions because such knowledge may lead to innovative strategies for snail-targeted schistosomiasis control. The studies have been focused on the model snail *B. glabrata* and four genomic regions involved in resistance have been reported<sup>23–26</sup>. Among them, three were identified from Guadeloupe and 13-16-R1 strains by association mapping or linkage disequilibrium (LD) mapping techniques<sup>24–26</sup>. An additional mapping effort applied genetic crosses between outbred M line and BS90 snails (which are different from the homozygous iM line and iBS90 snails used in the present study) and randomly amplified polymorphic DNA (RAPD)-based bulked segregant analysis (BSA)<sup>23</sup>. In the present study, we utilized F2 individuals derived from our two homozygous lines, ddRADseq genotyping, and LOD-based genetic analysis to identify two major QTLs, qRS-2.1 and qRS-5.1.

Our comprehensive genome-wide marker analysis has already demonstrated that qRS-2.1 is a large genomic locus (not a single





**Fig. 6** LOD profile of QTLs detected on linkage group 1 (LG1) to LG18 in *B. glabrata*. Two QTLs were detected on LG2 and LG5 for host resistance/susceptibility (in red) and parasite abundance (in blue), while one QTL detected on LG 13 for pigmentation (in green). Gray dot lines show threshold LOD score (LOD = 4) used for detecting QTL.

gene locus or marker) showing an overdominance effect for susceptibility (or underdominance effect for resistance), a genetic effect not often considered in *B. glabrata*. To further validate this mode of genetic effect, we investigated odds of infection based on the number of infected and not infected snails, as well as an F-test based on the density (read counts) of schistosome parasites present in the snail body in the SNP marker m2-37212557 of QTL qRS-2.1. The odds of infection for heterozygotes (AB) (susceptible snails (SS) = 45; resistance snails (RS) = 21) is significantly higher than that for either AA (susceptible alleles) homozygotes (SS = 10; RS = 16) or for BB (resistance alleles) homozygotes (SS = 5; RS = 19) ( $45/21 > 10/16$ ;  $P < 0.05$ ;  $45/21 > 5/19$ ;  $P < 0.05$ ) (detailed data provided in Supplementary Data 2). Furthermore, we have analyzed the density of parasites in snails with different genotypes (Supplementary Data 2) using density/abundance of parasite reads present in the snail body which was confirmed as an excellent phenotype of resistance/susceptibility for QTL mapping (see Fig. 6). Apparent overdominance was shown based on calculations showing that individuals with heterozygous combinations of alleles had significantly higher parasite read counts (i.e., more parasites in snail body) than either of the homozygous combinations of the alleles at the marker m2-

37212557 (mean parasite density AA =  $0.21 \pm 0.05$ , AB =  $0.35 \pm 0.03$ , BB =  $0.12 \pm 0.05$ ,  $F_{2,113} = 8.87$ ,  $P < 0.001$ ). Together, the two independent tests focused on the marker m2-37212557 as presented here suggested a mode of genetic effect of overdominance for susceptibility (or underdominance for resistance), which is in agreement with our conclusion based on genome-wide Simple Interval Mapping described in the Results section.

Among all 18 LGs, qRS-5.1 showed the highest LOD profile and like the other previously reported QTLs in *B. glabrata*, is associated with resistance to schistosomes (additive effect of resistance). In addition, we also noted loci with high LOD scores (but  $< 4$ ) in LG4 and LG10. LG10 was found to harbor the previously identified RADres resistance locus<sup>25</sup>. It is intriguing that our two most prominent QTLs (qRS-2.1 and qRS-5.1) are different from, and are even on different chromosomes than, the QTLs identified by the previous studies (Fig. 7). Several plausible reasons may explain the differences. First, genetic background and allelic effects may differ among strains or cross populations (i.e., M line/BS90 crosses<sup>23</sup>, five-generation selected snails that originated from Guadeloupe<sup>24</sup>; 13-16-R1 strain<sup>25,26</sup>, F2 snails bred from iM line/iBS90 in current study), due to interactions between parental alleles and genetic background<sup>47</sup>. Second,

**Table 2 Quantitative trait loci (QTL) of schistosome-resistance and pigmentation detected by simple interval mapping (SIM).**

Trait	QTL	LG	Position (95% CI) (cM)	QTL peak Left Marker	QTL peak Right Marker	LOD	PVE (%)	Add	Dom	QTL peak size (95% CI size) (cM)	QTL peak size (95% CI size) (MB)
Resistance	qRS-2.1	2	62 (61.5, 66.5)	m2-36225505	m2-37212557	4.47	12.73	0.06	0.41	1.43 (5.0)	0.99 (7.95)
	qRS-5.1	5	74 (72.5, 79.5)	m8-12003253	m8-13368092	4.79	13.27	0.30	0.03	0.50 (7.0)	1.36 (6.13)
Parasite density	qsm-2.1	2	62 (61.5, 63.5)	m2-36225505	m2-37212557	4.02	11.60	0.03	0.20	1.43 (2.0)	0.99 (4.03)
	qsm-5.1	5	74 (71.5, 79.5)	m8-12003253	m8-13368092	4.92	13.70	0.16	-0.01	0.50 (8.0)	1.36 (21.36)
Pigmentation	qBC-13.1	13	85 (84.5, 85.5)	m49-1163967	m49-1796291	8.92	63.20	0.34	-0.18	0.24 (1.0)	0.63 (0.89)

LG linkage group. cM CentiMorgan. LOD logarithm of the odds. CI confidence interval. PVE percentage of total phenotypic variance explained by the QTL. Add additive effect. Dom dominance effect.

interaction between host and parasite may be affected by genotypes of both host and parasite, and the schistosome strains used in previous studies and the present study differ<sup>24,25</sup>. Third, the degree of associations between a genetic marker and resistance trait in natural populations results from historical recombination which may vary among populations, whereas linkage mapping infers recombination event in populations derived from control crosses between phenotypically divergent strains<sup>48,49</sup>.

The two QTLs we identified are in the low-recombining centromeric and pericentromeric regions of LG2 and LG5 (Supplementary Fig. 8). The clustering of QTLs in the centromeric and pericentromeric regions has also been observed in many plants, such as root-related traits of wheat<sup>50</sup>, grain weight of rice and wheat<sup>51,52</sup>, and head blight resistance of wheat<sup>53</sup>. In *Drosophila*, a QTL affecting knockdown resistance to high temperature in *D. melanogaster* and a QTL for *D. sechellia* resistance to toxins were found to be strongly associated with the pericentromeric regions of the third chromosome<sup>54,55</sup>. In *B. glabrata*, two resistance loci, GRC/SOD (in LG6) and RADres (in LG10), were found in a low-recombining centromeric/pericentromeric region<sup>24,25</sup>. As noted by Tennessen et al.<sup>25</sup>, this region consists of a large haplotype block showing little recombination and containing several immune-relevant genes. Therefore, these haplotype blocks in the regions of low recombination may represent the combined effect of multiple contributing loci<sup>56</sup>. Gene ontology (GO) enrichment analysis of candidate genes within the 95% confidence intervals of QTL regions indicated that several functional candidate genes are involved in binding functions (RNA-binding, DNA binding, ATP binding, and zinc finger Ran-binding), cellular and membrane function (transmembrane and epithelial membrane). This finding is partially consistent with recent studies suggesting that transmembrane proteins play a role in resistance to infections in *B. glabrata*<sup>24,26</sup>. Although most of genes identified in QTLs are uncharacterized or unknown in defense, we examined 49 genes from these two QTLs and found 28 are significantly upregulated in BS90 snails at 12 h post exposure to *S. mansoni*<sup>57</sup>. Functional investigations (e.g., RNAi-knockdown and/or CRISPR-mediated knockout) of these genes, especially those located at the peak of LODs, are needed.

In addition to the resistance/susceptibility loci, we revealed a locus that influences snail pigmentation. In this locus, we identified a gene which is homologous to the G protein-coupled receptor NIA42. This gene has been found to play an essential role in melanization of the brown planthopper *Nilaparvata lugens*<sup>58</sup>, but its role in molluscan species is unknown. Identifying QTL and genes that are involved in pigmentation has not yet been reported in the Mollusca, including schistosome vector snails.

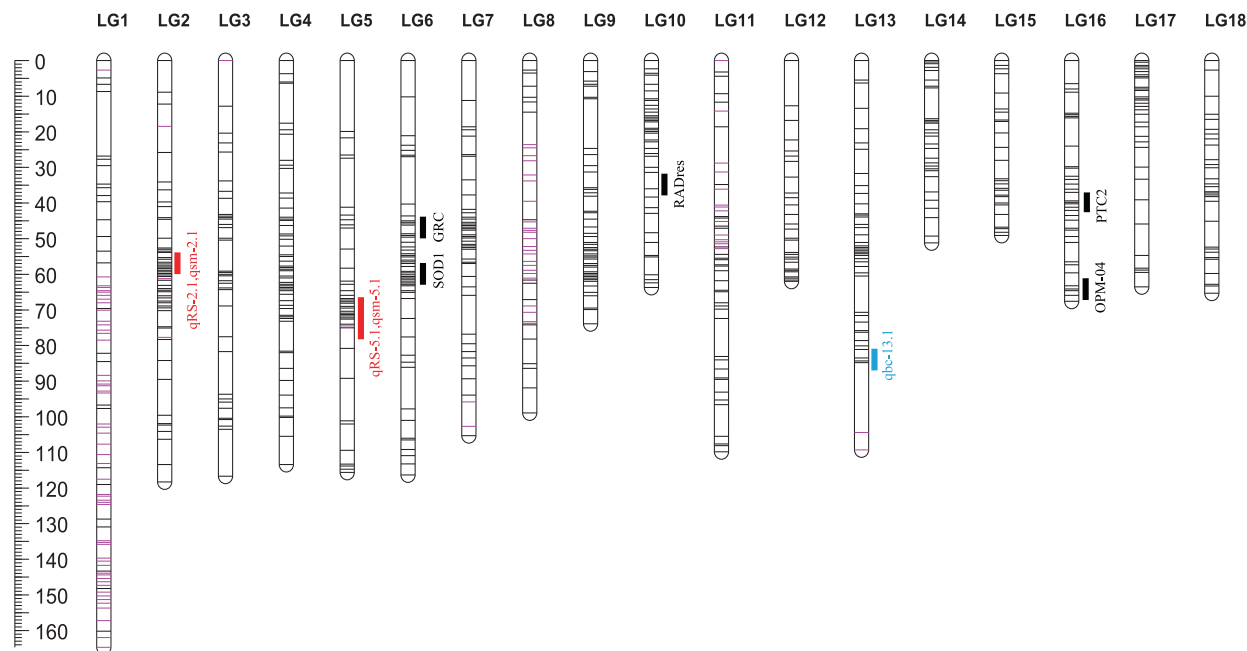
As this study utilized our newly developed genetic and genomic resources, two independent phenotype data (cercarial shedding and abundance of schistosome reads), and powerful ddRADseq-based genotyping, we believe our data are highly reliable. As resistance is controlled by multiple loci, it is highly likely that other loci may also be involved in resistance<sup>31,32</sup>. The effect of different loci on schistosome resistance may vary, likely depending on the snail or schistosome strains.

We are also aware of some potential limitations for genetic mapping presented in this study. First, the QTLs were mapped to large genomic regions in this study due to low map resolution resulting from the limited number of crossovers of the F2 population. Further effort will be made by applying advanced genetic resources such as recombinant inbred lines (RILs) or lines from an advanced intercross because these genetic resources will increase mapping resolution through the accumulation of additional meiotic crossover events. Second, a relatively small population size of 116 F2 snails in this study may not have provided

**Table 3 List of some key loci/genes in the 18 linkage groups.**

Linkage group	G3-LG	Genetic distance (cM)	Physical distance (MB)	Key loci/genes	References
LG1	I	164.7	90.5	inPhox, Ferritin, duox1, duox303, duox584, Matrilin1, BgGRN, ITGA3, prx621	101-103
LG2	II	118.3	66.7	BgCREB, Coagulation factor XI, qRS-2.1, BgRelish	101,104
LG3	III	116.7	63.6	P450If2, P450-2C3, BgRel, GalectinIIa	101,104-106
LG4	IV	113.4	55.4	PGRP-SA, PGRP, PGRP-LA1, PGRP-LA2	107
LG5	V	115.6	53.6	prx1, TPx, gp65, HSP40, qRS-5.1, gp97, GalectinIIa, FReM, Cystatin	101,102,105,108, this study
LG6	IX	116.3	53.6	GRC, cat42, BgTLR, spondin-1, bmplys1, bmplys2, bmplys3, SOD1, prx4, aif, nox2, schistosomin	24,25,101,102,109-113
LG7	VI	105.3	49.4	LBP, phox22, GGBP, $\beta$ -1, 3-glucanase, P4503A24, Inter alpha-trypsinin inhibitor	101,102,105,107
LG8	VIII	98.9	47.2	hcl-1, hcl-2, BgTEP, bmplys 8-12, 15, 17-23, P4503A56	105,112,114,115
LG9	XI	73.9	44.6	fibrillin-1, importin-7	This study
LG10	VII	63.7	43.8	BgSTAT1, BgSTAT2, RADres	25,104
LG11	XVIII	109.8	43.4	AIG-a, -c, -e, -h, -i, -j, -k	75
LG12	XIII	62.0	41.0	Theromacin, gp2404, Matrilin, MIF2	101,102
LG13	X	109.3	37.1	p63, p53, HSP20, FREP1, FREP12, FREP3, FREP14, FREP11, FREP2, FREP6, FREP13, FREP7, FREP5, FREP10, FREP4, qBC-13.1	101,108,116-120, this study
LG14	XV	51.2	34.3	FnBPA, acetylcholinesterase, BgTLR3, BgTLR13, catalase	This study
LG15	XIV	49.1	31.5	Metalloproteinase, AIG-f, bmplys13	75,101,112
LG16	XII	67.6	29.7	Dermatopontin, MIF1, bmplys4-7, HSP70B2, PTC2, AIG-g, OPM-04	23,26,75,102,105,112,121
LG17	XVI	63.5	29.1	arg1386, arginase	102,106
LG18	XVII	65.3	24.5	$\beta$ -IAP, SPI, HSP70	101,105

G3-LG: LGs reported by Tennesen et al.<sup>43</sup>. Please also see Supplementary Table 11 for the comparison between the two linkage maps.



**Fig. 7 Distribution of QTLs and SNP markers on the 18 LGs.** The markers derived from segregating normal and segregating distorted regions are presented as black and pink, respectively. Red solid bars indicate the QTL regions of snail resistance to parasites, and the blue solid bar show the QTL region for pigmentation identified in this study. The black solid bars indicate the QTL regions reported previously. OPM-04<sup>23</sup>, GRC<sup>24</sup>, RADres and SOD<sup>25</sup>, PTC<sup>26</sup>. The RAPD marker OPZ-11 is not indicated because it is a repetitive sequence which could not be unambiguously mapped.

enough power to detect QTLs with modest effect. In this study, the power of QTL detection within the 5 cM support interval was around 80% at a Type I error rate  $\alpha = 0.01$  to explain 10% of the phenotypic variance in the 116 F2 individuals, based on the calculation of the statistical power for detecting a QTL located in a marker interval<sup>59</sup>. Clearly, a large sample size will further improve power for QTL detection. Third, testing offspring that are bred from parental strains may also affect the outcome of QTL assignment. Since parental strains are unlikely to contain segregating alleles of large effect at every locus contributing to variation in the mapping population, some novel QTLs will remain undetected if the mapping population contains no segregating alleles at the loci. Use of population-based genome-wide association studies (GWAS) may be an option<sup>60,61</sup>.

In conclusion, this study selected two homozygous lines of the snail *B. glabrata* with contrasting resistance phenotype, generated two high-quality genome sequences, provided high-coverage linkage map, and identified new QTLs involved in schistosome resistance and pigmentation in snails. These genetic and genomic resources lay a foundation for further molecular studies of snail-schistosome interactions and yield new insights into mechanisms of snail resistance/susceptibility to schistosomes, which may ultimately benefit the development of innovative strategies for schistosomiasis control.

## Methods

**Ethics statement.** Mice were used to maintain the schistosome life cycles and to produce miracidia for schistosome-exposure experiments. Infection is through contact with cercariae-containing water and involves minimal discomfort. Infected mice are euthanized with CO<sub>2</sub> prior to showing clinical signs of disease and are dissected to recover parasitic eggs. The animal facilities and operations are maintained in compliance with the Animal Welfare Act regulations of the U.S. Department of Agriculture and guidelines for Public Health Service and Use of Laboratory Animals. The use of the mice was approved by the University of New Mexico Institutional Animal Care and Use Committee (IACUC protocol # 19-200895-MC).

**Breeding the two homozygous snail lines.** The new lines, iM line and iBS90, described in the paper were bred from *B. glabrata* M line and BS90, respectively.

Breeding the new selfed M line and BS-90 snail lines was started in 2001 and 2011 in our laboratory, respectively. Producing such highly inbred lines has proven difficult in many species because inbreeding depression results in a cessation of egg-laying. With careful maintenance, our homozygous snails are able to survive and reproduce. To avoid losing this unique resource, we stopped selfing the iM line at 81 generations of consecutive self-fertilization, in December 2018. Since then, this genetically stable new iM line has been well-established in our laboratory. We have continued breeding new iBS90 snails by selfing, which have now undergone 39 consecutive generations of self-fertilization. We have selected and tested the resistance/susceptible phenotype of either line from the beginning of the breeding experiment. For iM line, 10 juvenile snails (at 3–5 mm of shell diameter) were exposed to *S. mansoni* at a dose of 20 miracidia per snails in every generation at the first 20 generations and after that, the same test/selection was done in every 5 generations. The same protocol was also applied to iBS90, except more frequently. At the first 20 generations, every generation was tested and selected. From 20 generations on, the test was conducted every two generations. The two well-established genetic lines were not only used for genome sequencing, but also for production of F2 offspring for subsequent genetic mapping. Supplementary Fig. 1a provides the scheme for breeding of the two lines as well as producing F2 snails for linkage and genetic mapping.

## Generating F2 snails and testing their phenotype of schistosome-resistance.

*Biomphalaria glabrata*, a hermaphroditic species, can self-fertilize, but is a preferential out-croser. To make sure the F2 offspring was produced by crossing of the two parent snails (iM line (78-generations of consecutive selfing) and iBS90 (28-generations of consecutive selfing) and not by self-fertilization of either parent, the following procedures were used. Single iM line and iBS90 snail were placed in a plastic cup (1L) and allowed to produce F1 progeny. As pigmentation is a dominant Mendelian trait<sup>60</sup>, three possible outcomes are expected in F1 snails; albino F1 snails produced by selfing of M line snails, pigmented F1 snails derived from selfing of iBS90, and pigmented F1 derived from a cross between iM line and iBS90. Albino F1 snails were discarded and pigmented F1 snails were kept for producing F2s by selfing. To separate the two types of pigmented F1 snails, we examined F2 snails. If all F2 snails derived from one F1 by selfing were pigmented, these F2 snails were not used for genetic studies because they were from an F1 snail that resulted from self-fertilization of a parental iBS90 snail. If F2 snails were mixed with albino and pigmented ones, it indicated that their parental F1 snail was produced from a cross between iM line and iBS90 parents. Accordingly, these F2 snails were kept for phenotype and genotype tests.

The F2 juveniles (3–5 mm shell diameter) were exposed individually to 15–20 miracidia of the PR1 strain of *Schistosoma mansoni*. Since cercariae play a decisive role in infection and transmission to the definitive hosts, we define the resistance or susceptible phenotype here as to whether the schistosome-exposed snails shed cercariae. Shedding of cercariae was conducted at 45 days post-exposure (DPE). To determine whether they were truly resistant (some snails may shed later),



additional shedding was conducted at 55 DPE. The snails that did not shed cercariae at 55 DPE were considered as resistance snails whereas snails that shed cercariae at any time were recorded as susceptible ones. After screening for the cercarial shedding, each F2 snail labeled with body color (pigmented or albino) and phenotype (resistance or susceptible) was placed individually in a 1.5 ml tube with CTAB solution<sup>62</sup> and then stored at  $-80^{\circ}\text{C}$ .

**Snail lines used.** Only two specimens for iM line and two for iBS90 were used for whole genome sequencing; one was sequenced using Illumina technology the other was done by PacBio Sequel II. For Illumina sequencing, the single iM line or iBS90 snail was collected at 74 or 18 generations of consecutive self-fertilization, respectively. For PacBio sequencing, snails from 81 or 30 generations of self-fertilization were collected from the two lines (Supplementary Fig. 1a).

**Illumina sequencing.** Genomic DNA was extracted from a single iM line or iBS90 snail (the intestine excluded) using CTAB<sup>62</sup>. DNA quality was assessed by agarose gel (1.0%) electrophoresis. Concentration and purity of the extracted DNA were evaluated using a Nanodrop 2000 spectrophotometer and a Qubit 3.0 Fluorometer (ThermoFisher Scientific). After selection of size (400–700 bp) using Blue Pippin, a 150 nt x 2 paired-end library was prepared (KAPA Hyper Prep Kit Illumina platforms, KAPA Biosystems, [www.kapabiosystems.com](http://www.kapabiosystems.com)) and sequenced on the Illumina NextSeq500 platform at the University New Mexico (UNM) Biology Department's Molecular Biology Facility (<http://ceti.unm.edu/core-facilities/molecular-biology.html>).

**PacBio sequencing.** One specimen from either iM line or iBS90 was used. DNA extraction was done using the Qiagen Genomic-tip kit ([www.qiagen.com](http://www.qiagen.com)). The snail was first dissected to remove the intestine, and then flash frozen in liquid nitrogen. The snail was ground in a mortar and pestle in liquid nitrogen and the powder was suspended in a solution of 30 mM Tris pH 8, 10 mM EDTA, 1% SDS with Proteinase K (Qiagen). This was incubated at  $50^{\circ}\text{C}$  for 30–60 min. DNA was isopropanol precipitated, washed in 80% ethanol, and then suspended in Qiagen genomic tip digestion buffer. The process was repeated as per Qiagen instructions, including the column cleanup. Library preparation was performed using the PacBio express kit 2.0 and instructions. Sequencing was performed following the PacBio Sequel II annealing/binding and cleanup instructions as found on SMRT link 7. PacBio Sequel II - SMRT Cell sequencing was conducted at Brigham Young University DNA Sequencing Center (<https://biology.byu.edu/dnasc>).

**Genome assembly.** The overall workflow for bioinformatics and genetic analysis is provided in Supplementary Fig. 9. To generate high-quality Illumina paired-end reads, Illumina raw data were trimmed and filtered using Trimmomatic v0.36<sup>63</sup>, with options of slide window of 4 nt, average base quality score 20 and minimum read length 36 nt cut-off. Decontamination of Illumina reads was performed using Kraken v2.0.8 with 8GB database of reference microorganisms<sup>64</sup>. PacBio sequences were corrected, trimmed and assembled to contigs using assembler Canu 1.9<sup>65</sup> with options (useGrid = true corOutCoverage = 100 genomeSize = 916 m). The assembled contigs were then corrected with cleaned Illumina paired-end sequence data using program Pilon 1.23<sup>66</sup>. The corrected contigs were further linked to scaffolds using SSPACE 3.0<sup>67</sup> based on the hint of clean Illumina paired-end sequences.

To check contaminations in initial assembly, modified workflow from Blobtools<sup>68</sup> were used. Scaffold sequences were split into 10 Kbp fragments using in house bash/perl scripts, which were compared against NCBI non-redundant nucleotide database using the BLASTn program in NCBI-BLAST + package 2.11.0<sup>69</sup> with options: -task megablast -outfmt '6 std qlen slen staxids stitle' -num\_threads 79 -max\_target\_seqs 1 -max\_hsps 1 -evaluate 1e-5 -perc\_identity 60. The BLAST output file was loaded into Blobtools to classify hits taxonomically at the phylum level for each scaffold. Scaffolds were marked as contaminated if phylum was assigned to bacterial genomes and found no hits to known Mollusca nucleotide sequences. Final scaffolds without strong enough signatures of contaminations were sorted based on length and renamed. Evaluation of core conservative genes in genome assembly were performed using Benchmarking Universal Single-Copy Orthologs BUSCO 4.1.4 with metazoan core gene dataset metazoa\_odo10<sup>30</sup>.

**Genome annotation.** Genome annotation for the two strains was conducted independently in three stages: (1) repeated mask to identify repeat elements; (2) gene model prediction in genomic regions outside of repeated regions; (3) functional annotation or assigning functions to predicted gene models. Repeats in genome were initially identified and classified using RepeatModeler 2.0.1<sup>70</sup>. The predicted repeat models were then checked to keep retrotransposon-related functional domains and to exclude any other functional domains predicted by InterProScan 5.45<sup>71</sup>. Clean repeat models were used to mask the genome using RepeatMasker 4.0<sup>72</sup>. Based on the repeats identified, the scaffolds were soft masked (repeats in lower-case letters) for gene prediction.

Gene models were predicted using Evidence Modeler 06/25/2012<sup>73</sup>, by integrating weighted evidence from ab initio predictions, RNA sequencing (RNA-Seq) alignments, and homology-based searches.

The ab initio prediction results were collected from the output of BUSCO, which identifies core single copy orthologs using customized HMM models built from tBLASTn<sup>69</sup> and AUGUSTUS predictions<sup>74</sup>. Annotation evidence was a collection of sequences merged from predicted transcripts of the *B. glabrata* BB02 genome, de novo assembled transcriptomes of tissue-specific RNA-Seq data<sup>27</sup>, and recent schistosome infection studies (RNAseq) on *B. glabrata* M line and BS90 strains<sup>75,76</sup>. Redundant transcripts were removed using CD-HIT-EST<sup>77</sup>. Protein sets were derived from the merged transcripts using EvidentialGene<sup>78</sup>. Sequences showing similarities (E value  $<10^{-5}$  and coverage  $>50\%$ ) to known repeat elements (RepeatPeps.lib in RepeatMasker libraries) were filtered out by BLASTp<sup>69</sup>.

Non-redundant transcript sequences generated above served as evidence in the Program to Assemble Spliced Alignments (PASA) pipeline<sup>79</sup>, which infers gene models consistent with the spliced alignments of expressed transcript sequences.

The RNA-Seq data from 12 tissues<sup>27</sup> and infection studies<sup>75,76</sup> were also used to aid the annotation. This involved sequence cleaned up using Trimmomatic v0.36 as described above in genome assembly, then alignment to the assembled genome using HISAT2<sup>80</sup>. The alignment files and evidence transcripts were used as input for the BRAKER 2.1.1 annotation pipeline<sup>81</sup>, in which GeneMark-ES<sup>82</sup> utilizes RNA-Seq reads to genome alignments and transcripts to genome alignments to generate annotation hints to train AUGUSTUS, in two rounds: first separately, then merged in a second round as described in the BRAKER2 pipeline.

For functional annotation, the coding sequences and protein sequences based on the predicted gene models were searched against reference databases: (1) BLASTp to the UniProt database<sup>83</sup> with minimum identity of 30%, max E value  $10^{-5}$ , and minimum aligned length 10 aa; (2) BLASTp to the NCBI non-redundant protein database (NR) with the same cut off above<sup>84</sup>; (3) BLASTn to the NCBI non-redundant nucleotide database (NT) with minimum identity at least 60%, maximum E value  $10^{-5}$ ; (4) conserved functional domain by InterProScan 5.45<sup>71</sup>, and *Biomphalaria* specific IgSF domains predicted by *B. glabrata* specific HMM model with minimum length of 40 aa, and max E value 0.001<sup>76</sup>.

**Genomic structural variation analysis.** To identify synteny and structural variations (SV) between iM line and iBS90 genomes, the draft genome of iBS90 was aligned as query using the iM line genome as reference, using minimap2 version 2.17-r941<sup>85</sup> with options (-ax asm5 -eqx). The scaffolds fell into 18 linkage groups of iM line and their corresponding syntenic scaffolds in iBS90 were selected for the following steps. The output alignments in SAM format were used to identify structural variations using SyRI<sup>86</sup> with options (-k -F S). Annotated genes overlapping with predicted SVs were filtered using bedtools v2.29.2<sup>87</sup>. Synteny and SVs results data were organized and visualized using R packages openxlsx\_4.1.5<sup>88</sup>, tidyverse\_1.3.0<sup>89</sup>, and ggplot2\_3.3.2<sup>90</sup>.

#### Double digest restriction-site associated DNA sequencing (ddRADseq).

Genomic DNA was extracted individually using CTAB and re-purified using the mollusk DNA kit ([www.zymoresearch.com](http://www.zymoresearch.com)). Quantity and quality of the DNA were checked using Nanodrop and Qubit, respectively. Genomic DNA was double-digested with the restriction enzymes, *EcoR* I and *Msp* I to prepare the ddRADseq libraries. The unique barcode of 8 nucleotide base pairs with an Illumina adaptor was assigned to each individual accession for tracing the samples. The adaptor-ligated DNA amplicons were pooled and DNA fragments of 400–600 bp length were selected for library constructions. The ddRADseq libraries were sequenced on the MiSeq Illumina platform using  $2 \times 150$  bp paired-end (PE) mode at Admera Health ([www.admerahealth.com](http://www.admerahealth.com)). A total of 122 F2 snails and 4 parent snails (2 iM line and 2 iBS90) were analyzed by the ddRADseq.

#### ddRADseq data processing and SNP genotyping.

Sequence reads were de-multiplexed by MiSeq Reporter software and assigned to sequenced individuals. Sequence quality was checked using FastQC v. 0.11.9<sup>91</sup>. Sequence trimming was performed using Trimmomatic v. 0.36<sup>63</sup> to filter poor quality reads, and to remove adapters, barcode primer as well as some universal primers. Reads were also trimmed if 10 bases within a sliding window fell below a Phred score of Q20, or if index barcode, enzyme-specific adapter or Illumina adapter sequences were present, indicating read-through beyond sequenced insert fragments. After reads cleaning, the high-quality reads with a minimum length of 36 bp of 126 samples were individually aligned to the two reference-genome iM line and BS90, respectively, by using the Burrows-Wheeler Alignment tool (BWA)<sup>92</sup> with default parameters. In addition, the 4 parents pooled reads were also aligned to the two reference genomes. The alignments were converted to Sequence Alignment Map (BAM) format and duplicated reads were removed with Samtools 1.2<sup>93</sup>.

To determine polymorphisms of SNP markers between resistant and susceptible parents, genomic regions of read mapping from four parents with coverage between 40 and 4000 reads were extracted using the 'Create Mapping Graph', 'Identify Graph Threshold Areas', and 'Calculus Track' tools in CLC Genomics Workbench 12.0 (QIAGEN). SNP calling was performed on the coverage-restricted regions for all four parent samples, using the 'Fixed ploidy variant detection' tool in the CLC. The VCF file obtained in CLC from read mapping of the parent samples was used as the variant track file.

The ‘Identify Known Mutations from mappings’ tool within the CLC was used to genotype the 122 F2 samples individually at the SNP loci in the VCF file. The following parameters were used to filter variants: a minimum coverage of 5, minimum frequency of 20%, Ignore non-specific matches = Read, Ignore broken pairs = No, Include partially covering reads = No. The SNP allele specific to the resistant or susceptible parent, in which parents must be homozygous and have different genotypes, and those segregating in F2 population were considered genetically informative for genetic linkage mapping construction. SNP allele frequency ranging from 20% to 80% were classified as heterozygous while those <20% and >80% were classified as homozygous alleles derived from a resistant parent and a susceptible parent, respectively<sup>94</sup>. The genotype data were converted from a nucleotide-based format to a parent-based format as ABH in IciMapping. The threshold of maximum genotype missing rate was set to 25%. SNPs with missing genotypes in either parent were also excluded from further analysis. The markers were named based on the scaffold # and the physical position of the SNP.

**Linkage map construction and QTL analysis.** To construct the genetic linkage map, PLINK<sup>95</sup> and Haploview<sup>96</sup> were used to analyze linkage disequilibrium (LD) and select tag SNPs that represent the genome-wide markers. Pearson’s Chi-squared test was performed to examine the goodness of fit to the expected 1:1 segregation ratio ( $P < 0.01$ ) of alleles and 1:2:1 ratio of genotypes segregating in the F2 population. The MAP functionality in QTL IciMapping version 4.2.53<sup>97</sup> was used to construct the linkage map: SNP markers were grouped using a logarithm of the odds (LOD) threshold of 8.0 and a recombination frequency of 0.14. Linkage groups (LGs) with 10 or less markers were deemed to be unlinked and dropped from further construction. Marker ordering within group was based on their physical position. Recombination frequency was determined with window size of 5. Recombination frequencies between markers were converted into centiMorgans using the Kosambi mapping function<sup>98</sup>. The best marker order with the shortest linkage map distance was selected, and a linkage map was constructed with the information regarding marker segregation based on chi-square goodness of fit. The software MapChart<sup>99</sup> was used to display the marker names and genetic distance on the linkage map. More unassigned scaffolds were added to linkage groups if the visualization of synteny between iM line and iBS90 scaffolds suggested the presence of a continuity bridge.

The R/qtl and R/ASMap packages in R Statistical Computing Environment were used to calculate recombination fractions and LOD scores for all marker pairs. The function pull.rf() in R/qtl package was used to pull out the recombination fractions and LOD scores as matrices. A plot of the LOD scores against the estimated recombination fractions for all marker pairs was made to inspect any problem markers that had higher recombination fraction values and higher LOD scores. The function heatMap in R/ASMap package was used to create a heat map for the estimated recombinant fractions and the LOD scores.

QTL detections were performed using BIP functionality in QTL IciMapping. The functions for conventional simple interval mapping (SIM) and inclusive composite interval mapping were selected to detect additive QTL and dominance QTL. The mapping parameter step was set at 1.0 cM, and the LOD threshold for QTL declarations was determined with a 1000 permutation test at a 95% confidence interval, which controls the number of false positives, known as Type I error<sup>36</sup>. A LOD support interval of 1.0-LOD from the peak LOD was used to define QTL interval with 95% coverage<sup>100</sup>. QTLs were designated according to standard nomenclature based on the trait name and chromosome location. For example, qRS-5.1 is the first QTL for snail resistance to *S. mansoni* on chromosome 5. A QTL was defined as a major QTL if percentage of variation explained (PVE) was >10% or a minor QTL if <10%.

**Marker validation.** To confirm genetic effects of the major QTL qRS-5.1, 12 out of the top 20 SNP markers (based on  $-\log_{10} p$ -value) located within the QTL regions were randomly selected for validation of genotypes determined by ddRADseq assay. Technical validation was performed by Sanger sequencing of 46 resistant (R) F2 snails (27 pigmentation (B) and 19 albino (W)), and 48 susceptible (S) F2 snails (29 pigmentation and 19 albino) that were phenotyped and used for ddRADseq, whereas biological validation was conducted by Sanger sequencing of an additional set of 48 randomly selected F2 snails (12RB, 12RW, 12SB, and 12SW) that were phenotyped and not used for ddRADseq. The RAPD marker OPM-04 (Genbank accession number AF078109) was also included as a comparison for allele effects on parasite infections. Primers were designed based on the flanking regions of the SNPs using Primer 3 software. List of the forward and reverse PCR primers is shown in Supplementary Table 4.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The raw sequences for genome assembly and ddRADseq, as well as assembled annotated genome sequences, were all submitted to NCBI project with accession number PRJNA769727.

Received: 12 January 2022; Accepted: 15 August 2022;

Published online: 09 September 2022

## References

- World Health Organization (WHO). *Schistosomiasis Fact Sheet* (2021).
- Seubert, J., Pohlke, R. & Loebich, F. Synthesis and properties of Praziquantel, a novel broad spectrum anthelmintic with excellent activity against Schistosomes and Cestodes. *Experientia* **33**, 1036–1037 (1977).
- Woodhall, D. M. et al. Use of geospatial modeling to predict *Schistosoma mansoni* prevalence in Nyanza Province, Kenya. *PLoS ONE* **8**, e71635 (2013).
- Shiff, C. Why reinvent the wheel? Lessons in schistosomiasis control from the past. *PLoS Negl. Trop. Dis.* **11**, e0005812 (2017).
- Zacharia, A., Mushi, V. & Makene, T. A systematic review and meta-analysis on the rate of human schistosomiasis reinfection. *PLoS ONE* **15**, e0243224 (2020).
- King, C. H., Sutherland, L. J. & Bertsch, D. Systematic review and meta-analysis of the impact of chemical-based mollusciciding for control of *Schistosoma mansoni* and *S. haematobium* transmission. *PLoS Negl. Trop. Dis.* **9**, e0004290 (2015).
- Sokolow, S. H. et al. Global assessment of schistosomiasis control over the past century shows targeting the snail intermediate host works best. *PLoS Negl. Trop. Dis.* **10**, e0004794 (2016).
- Sokolow, S. H. et al. To reduce the global burden of human schistosomiasis, use ‘old fashioned’ snail control. *Trends Parasitol.* **34**, 23–40 (2018).
- Hubendick, B. A possible method of schistosome-vector control by competition between resistant and susceptible strains. *Bull. World Health Organ* **18**, 1113–1116 (1958).
- Richards, C. S. Genetics of a molluscan vector of schistosomiasis. *Nature* **227**, 806–810 (1970).
- Adolfi, A. et al. Efficient population modification gene-drive rescue system in the malaria mosquito *Anopheles stephensi*. *Nat. Commun.* **11**, 5553 (2020).
- Nolan, T. Control of malaria-transmitting mosquitoes using gene drives. *Philos. Trans. R. Soc. B Biol. Sci.* **376**, 20190803 (2021).
- Coelho, P. M. Z. et al. Transmission control of schistosomiasis mansoni by introduction of a resistant strain of *Biomphalaria tenagophila* in areas where transmission is maintained by this species. *Acta Trop.* **108**, 245–248 (2008).
- Marques, D. P. D. A. et al. Reduced susceptibility of a *Biomphalaria tenagophila* population to *Schistosoma mansoni* after introducing the resistant Taim/RS strain of *B. tenagophila* into Herivelton Martins stream. *PLoS ONE* **9**, e99573 (2014).
- Famakinde, D. O. Treading the path towards genetic control of snail resistance to schistosome infection. *Trop. Med. Infect. Dis.* **3**, 86 (2018).
- Maier, T. et al. Gene drives for schistosomiasis transmission control. *PLoS Negl. Trop. Dis.* **13**, e0007833 (2019).
- Newton, W. L. The inheritance of susceptibility to infection with *Schistosoma mansoni* in *Australorbis glabratus*. *Exp. Parasitol.* **2**, 242–257 (1953).
- Loker, E. S. Gastropod immunobiology. *Adv. Exp. Med. Biol.* **708**, 17–43 (2010).
- Adema, C. M. et al. Will all scientists working on snails and the diseases they transmit please stand up? *PLoS Negl. Trop. Dis.* **6**, e1835 (2012).
- Knight, M. et al. Schistosomes and snails: a molecular encounter. *Front. Genet.* **5**, 230 (2014).
- Lewis, F. A., Liang, Y., Raghavan, N. & Knight, M. The NIH-NIAID schistosomiasis resource center. *PLoS Negl. Trop. Dis.* **2**, e267 (2008).
- Paraense, W. L. & Correa, L. R. Variation in susceptibility of populations of *Australorbis glabratus* to a strain of *Schistosoma mansoni*. *Rev. Inst. Med. Trop. Sao Paulo* **5**, 15–22 (1963).
- Knight, M. et al. The identification of markers segregating with resistance to *Schistosoma mansoni* infection in the snail *Biomphalaria glabrata*. *Proc. Natl Acad. Sci. USA* **96**, 1510–1515 (1999).
- Tennessen, J. A. et al. Hyperdiverse gene cluster in snail host conveys resistance to human schistosome parasites. *PLoS Genet.* **11**, e1005067 (2015).
- Tennessen, J. A. et al. Genome-wide scan and test of candidate genes in the snail *Biomphalaria glabrata* reveal new locus influencing resistance to *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **9**, e0004077 (2015).
- Tennessen, J. A. et al. Clusters of polymorphic transmembrane genes control resistance to schistosomes in snail vectors. *Elife* **9**, e59395 (2020).
- Adema, C. M. et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* **8**, 15451 (2017).
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).
- Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).

30. Seppey, M., Manni, M. & Zdobnov, E. M. In *Gene Prediction. Methods in Molecular Biology* (ed. Kollmar, M.) Vol. 1962, 227–245 (Springer, 2019).
31. Richards, C. S. & Shade, P. C. The genetic variation of compatibility in *Biomphalaria glabrata* and *Schistosoma mansoni*. *J. Parasitol.* **73**, 1146–1151 (1987).
32. Richards, C. S., Knight, M. & Lewis, F. A. Genetics of *Biomphalaria glabrata* and its effect on the outcome of *Schistosoma mansoni* infection. *Parasitol. Today* **8**, 171–174 (1992).
33. Raghunathan, L. The karyotype of *Biomphalaria glabrata*, the snail vector of *Schistosoma mansoni*. *Malacologia* **15**, 447–450 (1976).
34. Hussain, W. et al. Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Sci. Rep.* **7**, 16394 (2017).
35. Xie, L., Klein, P., Crosby, K. & Jifon, J. A genotyping-by-sequencing Single Nucleotide Polymorphism-based map and genetic analysis of root traits in an interspecific tomato population. *J. Am. Soc. Hortic. Sci.* **144**, 394–404 (2019).
36. Churchill, G. A. & Doerge, R. W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
37. Berriman, M. et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature* **460**, 352–358 (2009).
38. Hofstatter, P. G., Thangavel, G., Castellani, M. & Marques, A. Meiosis progression and recombination in holocentric plants: What is known? *Front. Plant Sci.* **12**, 658296 (2021).
39. Grohme, M. A. et al. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* **554**, 56–61 (2018).
40. Bai, B. et al. Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. *Sci. Rep.* **8**, 691 (2018).
41. Kajiya-Kanegae, H. et al. RAD-seq-based high-density linkage map construction and QTL mapping of biomass-related traits in sorghum using the Japanese landrace Takakibi NOG. *Plant Cell Physiol.* **61**, 1262–1272 (2020).
42. Konar, A. et al. High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. *BMC Genomics* **18**, 417 (2017).
43. Tennessen, J. A., Bollmann, S. R. & Blouin, M. S. A targeted capture linkage map anchors the genome of the schistosomiasis vector snail, *Biomphalaria glabrata*. *G3 Genes Genom. Genet.* **7**, 2353–2361 (2017).
44. Lyttle, T. W. Segregation distorters. *Annu. Rev. Genet.* **25**, 511–581 (1991).
45. Li, X., Wang, X., Wei, Y. & Brummer, E. C. Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. *Theor. Appl. Genet.* **123**, 667–679 (2011).
46. Dai, B., Guo, H., Huang, C., Ahmed, M. M. & Lin, Z. Identification and characterization of segregation distortion loci on cotton chromosome 18. *Front. Plant Sci.* **7**, 2037 (2017).
47. Kane, K. L. et al. Genetic background effects on age-related hearing loss associated with *Cdh23* variants in mice. *Hear. Res.* **283**, 80–88 (2012).
48. Sawyer, S. L. et al. Linkage disequilibrium patterns vary substantially among populations. *Eur. J. Hum. Genet.* **13**, 677–686 (2005).
49. Nantawan, U., Kanchana-udomkan, C., Bar, I. & Ford, R. Linkage mapping and quantitative trait loci analysis of sweetness and other fruit quality traits in papaya. *BMC Plant Biol.* **19**, 449 (2019).
50. Soriano, J. M. & Alvaro, F. Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. *Sci. Rep.* **9**, 10537 (2019).
51. Li, J., Thomson, M. & McCouch, S. R. Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. *Genetics* **168**, 2187–2195 (2004).
52. Guan, P. et al. Dissection and validation of a QTL cluster linked to *Rht-B1* locus controlling grain weight in common wheat (*Triticum aestivum* L.) using near-isogenic lines. *Theor. Appl. Genet.* **133**, 2639–2653 (2020).
53. Buerstmayr, M. et al. High-resolution mapping of the pericentromeric region on wheat chromosome arm 5AS harbouring the Fusarium head blight resistance QTL *Qfhs.ifa-5A*. *Plant Biotechnol. J.* **16**, 1046–1056 (2018).
54. Jones, C. D. The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* **149**, 1899–1908 (1998).
55. Norry, F. M., Dahlgard, J. & Loeschcke, V. Quantitative trait loci affecting knockdown resistance to high temperature in *Drosophila melanogaster*. *Mol. Ecol.* **13**, 3585–3594 (2004).
56. Noor, M. A. F., Cunningham, A. L. & Larkin, J. C. Consequences of recombination rate variation on quantitative trait locus mapping studies: simulations based on the *Drosophila melanogaster* genome. *Genetics* **159**, 581–588 (2001).
57. Lu, L., Bu, L., Zhang, S.-M., Buddenborg, S. K. & Loker, E. S. An overview of transcriptional responses of schistosome-susceptible (M line) or -resistant (BS-90) *Biomphalaria glabrata* exposed or not to *Schistosoma mansoni* infection. *Front. Immunol.* **12**, 805882 (2022).
58. Uchiyama, H., Maehara, S., Ohta, H., Seki, T. & Tanaka, Y. Elevenin regulates the body color through a G protein-coupled receptor NIA42 in the brown planthopper *Nilaparvata lugens*. *Gen. Comp. Endocrinol.* **258**, 33–38 (2018).
59. Hu, Z. & Xu, S. A simple method for calculating the statistical power for detecting a QTL located in a marker interval. *Heredit* **101**, 48–52 (2008).
60. Vendelbo, N. M. et al. Discovery of resistance genes in rye by targeted long-read sequencing and association genetics. *Cells* **11**, 1273 (2022).
61. Kim, Y. et al. Genome-wide association study of actinic keratosis identifies new susceptibility loci implicated in pigmentation and immune regulation pathways. *Commun. Biol.* **5**, 386 (2022).
62. Winpenneinckx, B., Backeljau, T. & De Wachter, R. Extraction of high molecular weight DNA from molluscs. *Trends Genet.* **9**, 407 (1993).
63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
64. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
65. Koren, S. et al. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
66. Walker, B. J. et al. An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
67. Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* **15**, 211 (2014).
68. Laetsch, D. R. & Blaxter, M. L. BlobTools: interrogation of genome assemblies. *F1000Research* **6**, 1287 (2017).
69. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
70. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
71. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
72. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0* (2015).
73. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
74. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntentically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637–644 (2008).
75. Lu, L., Loker, E. S., Zhang, S.-M., Buddenborg, S. K. & Bu, L. Genome-wide discovery, and computational and transcriptional characterization of an AIG gene family in the freshwater snail *Biomphalaria glabrata*, a vector for *Schistosoma mansoni*. *BMC Genomics* **21**, 190 (2020).
76. Lu, L., Loker, E. S., Adema, C. M., Zhang, S.-M. & Bu, L. Genomic and transcriptional analysis of genes containing fibrinogen and IgSF domains in the schistosome vector *Biomphalaria glabrata*, with emphasis on the differential responses of snails susceptible or resistant to *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **14**, e0008780 (2020).
77. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
78. Gilbert, D. Gene-omes built from mRNA-seq not genome DNA. 7th Annual Arthropod Genomics symposium. *Notre Dame. F1000Research* **5**, 1695 (2013).
79. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
80. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
81. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinform.* **3**, lqaa108 (2021).
82. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**, e119 (2014).
83. Bateman, A. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
84. Agarwala, R. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
85. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
86. Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
87. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
88. Schaubberger, P. & Walker, A. *openxlsx: Read, Write and Edit xlsx Files* (2020).
89. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
90. Wickham, H. *ggplot2: Elegant Graphics for Data analysis* (Springer, 2016).
91. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* (2010).
92. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).



93. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
94. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12**, 443–451 (2011).
95. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
96. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
97. Meng, L., Li, H., Zhang, L. & Wang, J. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **3**, 269–283 (2015).
98. Kosambi, D. D. The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172–175 (1943).
99. Voorrips, R. E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78 (2002).
100. Dupuis, J. & Siegmund, D. Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* **151**, 373–386 (1999).
101. Mitta, G. et al. Gene discovery and expression analysis of immune-relevant genes from *Biomphalaria glabrata* hemocytes. *Dev. Comp. Immunol.* **29**, 393–407 (2005).
102. Larson, M. K., Bender, R. C. & Bayne, C. J. Resistance of *Biomphalaria glabrata* 13-16-R1 snails to *Schistosoma mansoni* PR1 is a function of haemocyte abundance and constitutive levels of specific transcripts in haemocytes. *Int. J. Parasitol.* **44**, 343–353 (2014).
103. Pila, E. A. et al. Endogenous growth factor stimulation of hemocyte proliferation induces resistance to *Schistosoma mansoni* challenge in the snail host. *Proc. Natl Acad. Sci. USA* **113**, 5305–5310 (2016).
104. Zhang, S.-M. & Coultas, K. A. Identification and characterization of five transcription factors that are associated with evolutionarily conserved immune signaling pathways in the schistosome-transmitting snail *Biomphalaria glabrata*. *Mol. Immunol.* **48**, 1868–1881 (2011).
105. Zhang, S.-M., Buddenborg, S. K., Adema, C. M., Sullivan, J. T. & Loker, E. S. Altered gene expression in the schistosome-transmitting snail *Biomphalaria glabrata* following exposure to niclosamide, the active ingredient in the widely used molluscicide bayluscide. *PLoS Negl. Trop. Dis.* **9**, e0004131 (2015).
106. Zhang, S.-M., Loker, E. S. & Sullivan, J. T. Pathogen-associated molecular patterns activate expression of genes involved in cell proliferation, immunity and detoxification in the amebocyte-producing organ of the snail *Biomphalaria glabrata*. *Dev. Comp. Immunol.* **56**, 25–36 (2016).
107. Zhang, S.-M., Zeng, Y. & Loker, E. S. Characterization of immune genes from the schistosome host snail *Biomphalaria glabrata* that encode peptidoglycan recognition proteins and gram-negative bacteria binding protein. *Immunogenetics* **59**, 883–898 (2007).
108. Zhang, S.-M., Nian, H., Zeng, Y. & DeJong, R. J. Fibrinogen-bearing protein genes in the snail *Biomphalaria glabrata*: characterization of two novel genes and expression studies during ontogenesis and trematode infection. *Dev. Comp. Immunol.* **32**, 1119–1130 (2008).
109. Galinier, R. et al. Biomphalysin, a new  $\beta$  pore-forming toxin involved in *Biomphalaria glabrata* immune defense against *Schistosoma mansoni*. *PLoS Pathog.* **9**, e1003216 (2013).
110. Goodall, C. P., Bender, R. C., Brooks, J. K. & Bayne, C. J. *Biomphalaria glabrata* cytosolic copper/zinc superoxide dismutase (*SOD1*) gene: association of *SOD1* alleles with resistance/susceptibility to *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **147**, 207–210 (2006).
111. Pila, E. A., Tarrabain, M., Kabore, A. L. & Hanington, P. C. A novel Toll-like receptor (TLR) influences compatibility between the gastropod *Biomphalaria glabrata*, and the digenean trematode *Schistosoma mansoni*. *PLoS Pathog.* **12**, e1005513 (2016).
112. Pinaud, S. et al. New insights into biomphalysin gene family diversification in the vector snail *Biomphalaria glabrata*. *Front. Immunol.* **12**, 635131 (2021).
113. Zhang, S.-M., Nian, H., Wang, B., Loker, E. S. & Adema, C. M. Schistosomin from the snail *Biomphalaria glabrata*: expression studies suggest no involvement in trematode-mediated castration. *Mol. Biochem. Parasitol.* **165**, 79–86 (2009).
114. Peña, J. J. & Adema, C. M. The planorbid snail *Biomphalaria glabrata* expresses a hemocyanin-like sequence in the albumen gland. *PLoS ONE* **11**, e0168665 (2016).
115. Portet, A. et al. BgTEP: An antiprotease involved in innate immune sensing in *Biomphalaria glabrata*. *Front. Immunol.* **9**, 1206 (2018).
116. Baričević, A., Štifanić, M., Hamer, B. & Batel, R. p63 gene structure in the phylum mollusca. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* **186**, 51–58 (2015).
117. Léonard, P. M., Adema, C. M., Zhang, S.-M. & Loker, E. S. Structure of two *FREP* genes that combine IgSF and fibrinogen domains, with comments on diversity of the *FREP* gene family in the snail *Biomphalaria glabrata*. *Gene* **269**, 155–165 (2001).
118. Zhang, S.-M., Léonard, P. M., Adema, C. M. & Loker, E. S. Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. *Immunogenetics* **53**, 684–694 (2001).
119. Zhang, S.-M. & Loker, E. S. The *FREP* gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and *FREP* retrosequences. *Dev. Comp. Immunol.* **27**, 175–187 (2003).
120. Zhang, S.-M., Adema, C. M., Kepler, T. B. & Loker, E. S. Diversification of Ig superfamily genes in an invertebrate. *Science* **305**, 251–254 (2004).
121. Garcia, A. B. et al. Involvement of the cytokine MIF in the snail host immune response to the parasite *Schistosoma mansoni*. *PLoS Pathog.* **6**, e1001115 (2010).

## Acknowledgements

We thank three anonymous reviewers for their constructive comments. This work was supported by National Institutes of Health's (NIH) grants R21 AI132953 (S.-M.Z.) and R01 AI170587 (S.-M.Z.). The Illumina data for the two genome sequences were generated through a pilot project of the UNM Center for Evolutionary and Theoretical Immunology (CETI) supported by NIH COBRE award 5P30GM110907 (E.S.L.). We would like to thank UNM Biology Department's Molecular Biology Facility for Illumina Sequencing and UNM Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high-performance computing and large-scale storage resources used in this work.

## Author contributions

Conceptualization: L.B., D.Z., and S.-M.Z.; data curation: L.B. and D.Z.; investigation: L.B., D.Z., L.L., E.S.L., G.Y., and S.-M.Z.; methodology: L.B., D.Z., and S.-M.Z.; project administration: S.-M.Z.; funding acquisition: S.-M.Z.; generation of the two snail lines: S.-M.Z.; breeding and phenotype-test of F2 snails: S.-M.Z.; software: L.B. and D.Z.; validation: L.B., D.Z., and S.-M.Z.; writing—original draft: L.B., D.Z., and S.-M.Z.; writing—review & editing: L.B., D.Z., L.L., E.S.L., G.Y., and S.-M.Z.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03844-5>.

**Correspondence** and requests for materials should be addressed to Si-Ming Zhang.

**Peer review information** *Communications Biology* thanks Jun Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Nishith Gupta and Christina Karlsson Rosenthal.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022