Three genes involved in the oxidative burst are closely linked in the genome of the snail, *Biomphalaria glabrata* ♠

Michael S. Blouin a,⁎, Kaitlin M. Bonner a, Becky Cooper b,1, Vindhya Amarasinghe c,1, Ryan P. O’Donnell d,1, Christopher J. Bayne a

a Dept. Zoology, Oregon State University, Corvallis, OR 97331, USA
b 28 Caudron Road, Hornby, Christchurch 8042, New Zealand
c Dept. Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA
d 5370 228th Ave. S.E., Issaquah, WA 98029, USA

⁎ Present address.

1 Present address.

Available online 1 December 2012

Article history:
Received 24 July 2012
Received in revised form 15 October 2012
Accepted 17 October 2012
Available online 1 December 2012

Keywords:
Snail
Schistosoma mansoni
Genetics
Superoxide dismutase
Catalase
Peroxiredoxin

A R T I C L E  I N F O

Article history:
Received 24 July 2012
Received in revised form 15 October 2012
Accepted 17 October 2012
Available online 1 December 2012

A B S T R A C T

Allelic variation at the Cu–Zn superoxide dismutase (*SOD1*) locus has been shown to be associated with resistance of the snail, *Biomphalaria glabrata*, to infection by the trematode parasite, *Schistosoma mansoni*. *SOD1* catalyses the production of hydrogen peroxide, a known cytotoxic component of the oxidative burst used in defence against pathogens. In our laboratory population of *B. glabrata*, the most resistant allele at *SOD1* is over-expressed relative to the other two alleles. Because hydrogen peroxide also causes oxidative stress on host tissues, we hypothesised that over-expression of *SOD1* might be compensated by epistatic interactions with other loci involved in oxidation–reduction (redox) pathways. Catalase, peroxiredoxins and glutathione peroxidases all degrade hydrogen peroxide. We tested whether alleles at each of these loci were in linkage disequilibrium with *SOD1* in our population, as might be expected given strong epistatic selection. We found that *SOD1*, catalase (*CAT*) and a peroxiredoxin locus (*PRX4*) are in strong linkage disequilibrium in our population. We also found that these loci are tightly linked, within 1–2 cM of each other, which explains the high linkage disequilibrium. This result raises the possibility that there is a linked cluster of redox genes, and perhaps other defence-relevant genes, in the *B. glabrata* genome. Whether epistatic interactions for fitness actually exist among these loci still needs to be tested. However the close physical linkage among *SOD1*, *PRX4* and *CAT*, and subsequent high disequilibrium, makes such interactions a plausible hypothesis.

© 2012 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

1. Introduction

The snail, *Biomphalaria glabrata*, is an intermediate host for the human pathogen, *Schistosoma mansoni*. Resistance in *B. glabrata* to infection with *S. mansoni* is highly heritable in many laboratory and field populations (Richards and Shade, 1987; Shoukry et al., 1997; Webster and Woolhouse, 1998; Webster et al., 2004; Theron et al., 2008). One locus at which allelic variation associates with resistance to the parasite is copper–zinc superoxide dismutase (*SOD1*) (Goodall et al., 2004, 2006; Bonner et al., 2012). Increased hydrogen peroxide (*H₂O₂*) production and increased expression of *SOD1* were correlated with the difference in resistance between snails from the M-line strain and the more resistant 13-16-R1 strain (Bender et al., 2005; Bayne, 2009). Within strain 13-16-R1, there are three common alleles (intron sequence variants A, B and C) that associate additively with resistance (Fig. 1). Furthermore, the most protective allele, B, is constitutively over-expressed relative to alleles A and C (Bender et al., 2007).

*SOD1* is a ubiquitous protein involved in several cellular functions including signalling and immune response (Nappi and Ottaviani, 2000; Fink and Scandalios, 2002; Ramasarma, 2007; Abreu and Cabelli, 2010). *SOD1* catalyses the reduction of highly reactive superoxide (*O₂⁻*) to *H₂O₂*. *H₂O₂* is a known cytotoxic component of the oxidative burst, which is one of the defence mechanisms for parasite clearance in molluscs (Bayne, 2009; Hahn et al., 2001; Loker, 2010). When PR-1 strain schistosomes invade a 13-16-R1-strain snail, hemocytes surround the invading parasite and are thought to generate *H₂O₂* as part of the killing mechanism (Hahn et al., 2001; Bender et al., 2007; Bayne, 2009).

*H₂O₂* is one of several reactive oxygen species that are used by cells for various biological processes in addition to defence against infection (e.g. signalling, growth and development, programmed cell death; Dowling and Simmons, 2009). However, reactive...
oxygent species are a double-edged sword. Their effects are dose-dependent and so mismatch between oxidative and anti-oxidative pathways can exert oxidative stress on cells, including damage to DNA, RNA and proteins (Beckman and Ames, 1998). Therefore, if over-expression of the B allele at SOD1 is responsible for its protective effect against parasites, then that raises the question of whether there are fitness costs to having that allele in the absence of parasites (Dowling and Simmons, 2009; Monaghan et al., 2009).

We recently showed that under laboratory conditions there was no fitness cost to having the B allele versus the C allele in terms of survival and fecundity, and there was a slight advantage in terms of growth (Bonner et al., 2012). One explanation for this unexpected result is that over-expression of the B allele is somehow compensated by epistatic interactions with other loci involved in oxidation-reduction (redox) pathways. Candidate loci that might compensate epistatically for over-expression of SOD1 include those coding for catalase, peroxiredoxins and glutathione peroxidases because these enzymes catalyse the conversion of H2O2 into water and other less harmful byproducts. However, such a compensatory mechanism would require SOD1 to be in substantial linkage disequilibrium (LD) with the interacting loci, which could occur via close physical linkage or strong, recurring epistatic selection (Sinnervo and Svensson, 2002; Phillips, 2008).

Here we tested whether loci coding for Cu–Zn superoxide dismutase (SOD1, GenBank accession number AY505496.2), catalase (CAT, FC856064), peroxiredoxin 4 (PRX4; Knight et al., 2009) and glutathione peroxidase (GPX; DW474737) were in LD in the 13-16-R1 snail population and if so, whether that could be explained by physical linkage in the genome. We show that SOD1, PRX4 and CAT are in substantial LD above background levels and that they occur in a <2 cM cluster on the same chromosome. GPX was not linked to, nor was it in LD with, the other three loci.

2. Materials and methods

2.1. Study population

We used a population of the 13-16-R1 strain of Biomphalaria glabrata at Oregon State University (OSU), USA since the mid-1970s. The 13-16-R1 snail strain was reportedly created from one or more crosses between strains of snails isolated from Brazil and Puerto Rico. Biomphalaria glabrata is a facultative self-fertilising hermaphrodite such that snails will preferentially outcross when given access to a mate, but when isolated will usually reproduce through self-fertilisation (e.g. Vianey-Liaud and Dussart, 2002; and personal observations). For example, our laboratory population is in Hardy–Weinberg equilibrium for all the loci used in this study. This population has been maintained in the absence of selection pressure by S. mansoni.

Previous work on SOD1 in this population looked at sequence variation in an approximately 540 bp piece of intron 4 (Goodall et al., 2006). Three main sequence variants (alleles) are segregating at this locus (alleles A, B and C) in addition to a rare variant of allele C (allele D at ~1% allele frequency), which in this study was scored as a C allele (only two copies were observed out of 80 individuals and the results were almost identical to when the D allele was included in the analyses).

2.2. Loci used

We used eight previously published microsatellite loci (Table 1; Jones et al., 1999; Mavarez et al., 2000) to assess the background level of LD in the population. We assumed that none of these eight loci are closely linked. Size variation at microsatellites was scored on an ABI 3100 capillary sequencer. Allelic variation at intron 4 of SOD1 was not scored in this study, which was scored only two copies were observed out of 80 individuals and the results were almost identical to when the D allele was included in the analyses).

### Table 1

<table>
<thead>
<tr>
<th>Microsatellite loci</th>
<th>Sample set 1</th>
<th>Sample set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hexp</td>
<td>N</td>
</tr>
<tr>
<td>Redox loci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOD1</td>
<td>0.68</td>
<td>3</td>
</tr>
<tr>
<td>PRX4</td>
<td>0.66</td>
<td>3</td>
</tr>
<tr>
<td>CAT</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>GPX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsatellite loci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bgc8</td>
<td>0.45</td>
<td>2</td>
</tr>
<tr>
<td>bgc7</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>bgc4</td>
<td>0.11</td>
<td>2</td>
</tr>
<tr>
<td>bgc5</td>
<td>0.66</td>
<td>6</td>
</tr>
<tr>
<td>bgc8</td>
<td>0.29</td>
<td>2</td>
</tr>
<tr>
<td>bgc10</td>
<td>0.55</td>
<td>4</td>
</tr>
<tr>
<td>bgc15</td>
<td>0.38</td>
<td>2</td>
</tr>
<tr>
<td>sub1</td>
<td>0.74</td>
<td>4</td>
</tr>
</tbody>
</table>
SOD1, intron 2 of GPX and intron 2 of PRX4 was assessed via Sanger sequencing or by high-resolution melt analysis on an ABI Fast Real-Time PCR system. The three alleles from intron 6 of CAT differed substantially in size (490, 566 and 592 bp), and therefore were scored by size on an ABI capillary sequencer. Sequences for the SOD1 alleles are available in GenBank under accession numbers DQ239577–DQ239579 (from Goodall et al., 2006), and for the alleles at CAT and PRX4 under JX989079–JX989081 and JX989076–JX989078. Allele sequences for GPX are shown in Supplementary Fig. S1. Primers and PCR conditions for the microsatellite and redox loci used in this study can be found in Supplementary Tables S1 and S2.

2.3. Linkage disequilibrium

Allelic variation was scored at SOD1, PRX4, CAT and the eight microsatellites on a sample of 40 snails taken from the laboratory population in 2006 (Sample set 1). We obtained complete genotypes for all 11 loci in all 40 snails in that sample. There was not enough DNA remaining in many of those samples to also score GPX. Therefore, we scored the four redox loci (but not the microsatellites) in an additional sample of 40 snails collected in 2008 (Sample set 2). LD among loci was analysed separately for each family as implemented in GENETIX 4.05.

2.4. Linkage mapping of SOD1, PRX4, CAT and GPX

There were three alleles at each locus. These were designated as follows: SOD1 (A, B and C), PRX4 (1, 2 and 3), CAT (490, 566 and 592) and GPX (100, 200 and 300) (Tables 1 and 2). Two crosses were performed. Family 1 was informative for mapping all four loci (CC 11 592592 100300 × AC 21 490592 200300). Family 2 (AB 22 566592 200300 × AC 12 592592 200200) was informative for mapping all pairs of loci except GPX-PRX4 and CAT-PRX4. Two hundred offspring from each family were genotyped at all four loci, with the exception that we genotyped only 22 individuals from Family 1 for GPX. This was because after first genotyping Family 2, it was obvious that GPX was unlinked to the other loci. Therefore we used a subset of individuals to verify the pattern in Family 1.

3. Results

3.1. Allelic variation

No loci showed significant deviations from Hardy–Weinberg equilibrium in either sample set (Table 1). The eight microsatellites showed levels of variation similar to those of the redox loci, with an average of 3.1 alleles per locus (range 2–6) and average heterozygosity = 0.47 (range 0.11–0.74), versus three alleles per locus and average heterozygosity of 0.62 (range 0.40–0.68) at the redox loci (Table 1). Thus, the microsatellites were good comparators for the background level of LD in the population. Table 2 shows allele frequencies at each of the redox loci.

3.2. Linkage disequilibrium in the 13-16-R1 population

LD among the eight putatively unlinked microsatellite loci ranged from 0 to 0.25, averaging 0.13 (Fig. 2). Two pairs of microsatellites (ubg5 with ubg1, and ubg8 with ubg10) showed significant LD but not after correction for multiple testing (28 pairwise tests).

Table 2

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample set 1</td>
<td>Sample set 2</td>
</tr>
<tr>
<td>SOD1</td>
<td>A</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.325</td>
</tr>
<tr>
<td>PRX4</td>
<td>1</td>
<td>0.275</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.35</td>
</tr>
<tr>
<td>CAT</td>
<td>490</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>566</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>592</td>
<td>0.75</td>
</tr>
<tr>
<td>GPX</td>
<td>100</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 2. Frequency distribution of pairwise R values (multi-allelic measure of linkage disequilibrium) between loci in Sample sets 1 and 2 (histogram with 0.01 unit bins). Pairwise values between microsatellite loci are shown in black bars, and between redox loci in white bars. Values for each pair of redox loci are labelled. Note that values among Cu–Zn superoxide dismutase (SOD1), peroxiredoxin 4 (PRX4) and catalase (CAT) are much larger than the background level in this population described by the eight putatively-unlinked microsatellites.

\[ \text{SOD1, Cu–Zn superoxide dismutase; PRX4, peroxiredoxin 4; CAT, catalase; GPX, glutathione peroxidase.} \]

\[ \text{SAlele frequency differences between Sample sets 1 and 2 are not statistically significant (tested for each locus using the exact test for population differentiation in GENEPOP).} \]
In contrast, LD was very high and statistically significant between SOD1 and PRX4 (\(R = 0.58\) and 0.65 in Sample sets 1 and 2, respectively), as well as high and significant between CAT and SOD1 (\(R = 0.37\) and \(R = 0.40\)) and between CAT and PRX4 (\(R = 0.25\) and \(R = 0.42\)) (Fig. 2; Table 3). LD between GPX and each of the three other redox loci ranged from 0.13 to 0.15 (Fig. 2; Table 3). Relative to the background level of LD in the population (inferred from the microsatellites), there is very high LD between SOD1 and PRX4, and high LD between those two loci and CAT (Fig. 2). LD between GPX and the other three redox loci was in line with the background level among putatively unlinked microsatellite loci in this population.

3.3. Estimated map distances among the four redox loci

No loci showed deviation from expected Mendelian inheritance in either family. In Family 1 we observed no recombinants out of 200 individuals among SOD1, PRX4 and CAT, giving an exact one-sided 95% confidence interval (CI) for the pairwise recombination rate, \(r\), among those three loci of \(<0.018\). In Family 2 we observed three recombinants out of 200 between SOD1 and CAT (\(r = 0.015\); 95% CI 0.003–0.043), and one recombinant between SOD1 and PRX4 (\(r = 0.005\); 95% CI 0.0001–0.028). Combining results from both families shows four recombinants out of 400, giving an estimate of 1 cm over the entire haplotype block (95% CI of 0.003–0.025). Therefore, we conclude that SOD1, PRX4 and CAT are closely linked to each other in a chromosomal block that probably spans less than 2 cm. Both families showed free recombination between GPX and the other loci.

3.4. Inference of gene order

Because we observed no recombinants in the fully-informative cross (Family 1), we cannot infer which locus is in the middle. However, from the strong positive or negative pairwise correlations between alleles across loci observed in the LD analysis, it was obvious that alleles at the three linked loci fall into three haplotype blocks (Supplementary data S1 and Fig. S2). The strongest associations among alleles distinguish haplotype 3, B, 566 from haplotypes 1, C, 592/490 and 2, A, 592/490 (for alleles at PRX4, SOD1 and CAT from right to left). Here 592/490 means either allele 490 or 592 at CAT (allele 490 is rare (Table 2) so there are effectively only two alleles at that locus). Given the strong signal of three ancestral haplotypes, we can count the number of putative recombinant chromosomes currently segregating in our population (e.g. a CB 13 566566 genotype would imply a recombination event occurred between CAT and the other two loci). In sample sets 1 and 2 we observed a total of 11 putative recombinants between PRX4 and the other two loci, and 15 putative recombinants between CAT and the other two loci. There were no putative recombinants between SOD1 and the other two loci. Therefore, we conclude that SOD1 is in the middle.

4. Discussion

Here we observed strong LD among SOD1, PRX4 and CAT that can be explained by their close physical linkage in a 1–2 cM block of chromosome. The three clear haplotype blocks present in the 13–16-R1 snail population probably reflect its hybrid origin between Caribbean and South American populations.

We currently do not know what physical distance corresponds to 1–2 cm in the B. glabrata genome. Average recombination rates of 1–5 cM per megabase of DNA are typical in other organisms, albeit with substantial variation across chromosomal regions (e.g. Backström et al., 2010; Stevison and Noor, 2010; Hohenlohe et al., 2012). We recently completed a single nucleotide polymorphism (SNP)-based linkage map from an outcrossed pair of 13-16-R1 snails, from which we estimated a very rough average of 4–10 cM/Mb (unpublished data). Thus, given no other information, it seems reasonable to tentatively conclude that SOD1, PRX4 and CAT are within several hundred thousand bp of each other. These three loci are found on three different, relatively small contigs in the current draft of the B. glabrata genome (which is still quite fragmented). An improved assembly will be required to determine the actual positions of these loci relative to each other.

We also have 20 inbred lines from the 13–16-R1 population in which we typed the four redox loci and several thousand SNPs (unpublished data). Relative to LD between SNPs that are known to be within 1–2 cm of each other, the LD observed among SOD1, PRX4 and CAT was not significantly higher. Thus, tight linkage alone is a sufficient explanation for the high LD among the redox loci (i.e. retarded decay of LD that was generated by genetic drift and/or mixture disequilibrium). Therefore, it is not necessary to invoke epistatic selection to explain the high LD.

On the other hand, given we chose, a priori, to examine LD among CAT, PRX4, SOD1 and GPX based entirely on their proteins’ enzymatic function, it does seem very coincidental that three turn out to be so closely linked. After obtaining our results, we checked whether loci for catalase, superoxide dismutase and peroxiredoxins are closely linked in some other well-annotated genomes for which good information on recombination rates is available. No evidence was found for linkage in vertebrates (human, chicken, cattle, zebra fish) or two insects (Apis mellifera, Anopheles gambiae). However, the three loci are linked in the same order (SOD1 in the middle) in Drosophila melanogaster and Caenorhabditis elegans, although not as tightly linked as they appear to be in B. glabrata (39 cM across the entire interval in D. melanogaster and 24 cM in C. elegans; Hillier et al., 2007; Fiston-Lavier et al., 2010). This similarity could simply reflect the arrangement in the common ancestor of insects, nematodes and molluscs. It is also possible that there is some selective advantage to having these loci physically linked.

Mounting evidence suggests that the oxidative burst is an important component of defence in molluscs (Bayne, 2009; Green et al., 2009; Mourao et al., 2009; Loker, 2010; Mone et al., 2011). Thus it is worth considering the possibility that there is a linked cluster of redox genes, and perhaps other defence-relevant genes, in the B. glabrata genome (e.g. similar to the Plasmodium resistance island in Anopheles; Riehle et al., 2006). Frequency-dependent biotic interactions (e.g. host-parasite) are considered among the most likely to increase LD among epistatically interacting loci (Sinervo and Svensson, 2002). Selection for reduced recombination among such loci would reduce the recombinational load, as in traditional models of the evolution of co-adapted gene complexes (Dobzhan-
the entire haplotype block that is driving the association with SOD1 closely-linked locus that actually controls resistance. More likely, it is possible that SOD1 is simply in strong LD with a closely-linked locus that actually controls resistance. It is possible that SOD1 makes such interactions a plausible hypothesis.

Our original motivation for testing whether the loci that control degradation of H₂O₂ were in LD with SOD1 was to explore the possibility that any costs of carrying the B allele at SOD1 are compensated by other loci (see Bonner et al., 2012). Whether such epistatic interactions for fitness actually exist still needs to be tested. However, the close physical linkage and high LD among SOD1, PRX4 and CAT makes such interactions a plausible hypothesis.

Acknowledgements

Thank you to J. Tennesen for assistance with the bioinformatics and SNP analysis, and to J. Tennesen and M. Steinauer for comments on an earlier draft. C. Goodall and R. Bender identified the PRX4 alleles and developed the high resolution melt assay. The Center for Genome Research and Biocomputing at Oregon State University (USA) assisted with genotyping. This work was partially funded by the US National Institutes of Health Grant AI016137.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijpara.2012.10.020.

References


