Discordance between nuclear and mitochondrial genomes in sexual and asexual lineages of the freshwater snail Potamopyrgus antipodarum

DOROTA PACZESNIAK,*† JUKKA JOKELA,*† KATELYN LARKIN‡ and MAURINE NEIMAN‡

*Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland, †Department of Environmental Systems Science, ETH-Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland, ‡Department of Biology, University of Iowa, 143 Biology Building, Iowa City, IA 52242, USA

Abstract

The presence and extent of mitonuclear discordance in coexisting sexual and asexual lineages provides insight into 1) how and when asexual lineages emerged, and 2) the spatial and temporal scales at which the ecological and evolutionary processes influencing the evolution of sexual and asexual reproduction occur. Here, we used nuclear single-nucleotide polymorphism (SNP) markers and a mitochondrial gene to characterize phylogeographic structure and the extent of mitonuclear discordance in Potamopyrgus antipodarum. This New Zealand freshwater snail is often used to study the evolution and maintenance of sex because obligately sexual and obligately asexual individuals often coexist. While our data indicate that sexual and asexual P. antipodarum sampled from the same lake population are often genetically similar, suggesting recent origin of these asexuals from sympatric sexual P. antipodarum, we also found significantly more population structure in sexuals vs. asexuals. This latter result suggests that some asexual lineages originated in other lakes and/or in the relatively distant past. When comparing mitochondrial and nuclear population genetic structure, we discovered that one mitochondrial haplotype ('1A') was rare in sexuals, but common and widespread in asexuals. Haplotype 1A frequency and nuclear genetic diversity were not associated, suggesting that the commonness of this haplotype cannot be attributed entirely to genetic drift and pointing instead to a role for selection.

Keywords: mating systems, parthenogenetic, polyploidy, selective sweep, SNP

Received 18 February 2013; revision received 29 May 2013; accepted 5 June 2013

Introduction

Both mitochondrial and nuclear genomes carry genetic information that allows reconstruction of phylogeographic relationships. The phylogenetic information provided by these two genomes is often in agreement (Avise 1994), but it is increasingly evident that mitonuclear discordance, defined as a significant difference in the patterns of differentiation between these two marker types (Toews & Brelsford 2012), is also common (reviewed in: Funk & Omland 2003; Toews & Brelsford 2012). Research aimed at revealing the source of this mitonuclear discordance has implicated mechanisms such as incomplete lineage sorting, differential selection on nuclear and mitochondrial genomes, and asymmetric hybridization. The latter is linked to phenomena like sex-biased dispersal, human-facilitated movement and Wolbachia infection (which can cause mating incompatibilities between infected and uninfected hosts) (Toews & Brelsford 2012).

Many of these sources of mitonuclear discordance are connected to or directly caused by canonical sexual reproduction or by more unusual forms of genetic exchange (reviewed in: Schurko et al. 2009; Toews & Brelsford 2012). As such, the extent of mitonuclear discordance in asexuals can illuminate four areas of fundamental importance to researchers addressing why sexual reproduction is so common in nature: 1) how new asexual lineages are generated (Neiman & Schwander 2011),...
2) their rate of origin (Burt 2000; Neiman et al. 2009), 3) how long these lineages persist (Neiman et al. 2009) and 4) the extent to which asexual lineages might experience occasional genetic exchange (Schurko et al. 2009).

Here, we used nuclear single-nucleotide polymorphism (SNP) markers and mitochondrial sequence data to build a comprehensive picture of population genetic structure and evaluate whether mitonuclear discordance is present in sexual and asexual lineages of Potamopyrgus antipodarum. This New Zealand freshwater snail has been used as a model system for studying the maintenance and distribution of sexual reproduction in nature for over 25 years (e.g. Lively 1987; Jokela et al. 1997, 2009). Sexual P. antipodarum are diploid and asexual P. antipodarum are typically polyploid (3χ and >3χ) (Wallace 1992; Neiman et al. 2011, 2012), although there is some evidence for the presence of diploid asexuals in two New Zealand lakes (see Methods and Neiman et al. 2011). Both allozyme (Dybdahl & Lively 1995) and mitochondrial data (Neiman & Lively 2004; Neiman et al. 2011) indicate that unlike many asexuals (recently reviewed in: Kearney 2005; Neiman & Sch wander 2011), asexual P. antipodarum are derived from sexual conspecifics rather than of hybrid origin. This latter point is important because it means that direct or indirect consequences of hybridization cannot account for differences between coexisting sexual and asexual individuals or subpopulations (Maynard Smith 1978; Jokela et al. 1997; Kearney 2005).

The nuclear (Dybdahl & Lively 1995) and mitochondrial data sets (Neiman & Lively 2004; Neiman et al. 2011) that had been used to test hypotheses of hybrid vs. nonhybrid origin in P. antipodarum were also used to address whether asexual P. antipodarum, which often coexist with sexual P. antipodarum (Lively 1987), were of recent local origin (recently derived from sympatric sexuals) vs. nonlocal and/or nonrecent origin (derived from allopatric sexuals and/or from sympatric sexuals in the relatively distant past). Whether asexual organisms are recently derived from sympatric sexual relatives represents critical information for evaluating whether sexual populations are likely to face threats from ecologically and phenotypically similar asexual lineages (Case & Taper 1986).

Dybdahl & Lively (1995) genotyped P. antipodarum from four New Zealand lake populations at six allozyme loci and showed that asexual P. antipodarum were more closely related to sympatric sexuals than to allopatric sexuals or asexuals, suggesting local and recent origin of asexual lineages from sympatric sexual conspecifics. In contrast to these results, two geographically broader mitochondrial sequence-based phylogeographic studies found that a majority of asexual P. antipodarum sampled from many different lakes shared a single mitochondrial haplotype that was rare in sexuals (Neiman & Lively 2004; Neiman et al. 2011).

Uncovering the reason for the strikingly different patterns of genetic structure among asexual P. antipodarum revealed by nuclear (allozyme) and mtDNA data requires genotyping multiple sexual and asexual individuals from many populations at both nuclear and mitochondrial loci and then using these data to compare and contrast the population genetic structure that these markers reveal. The most straightforward potential outcome of this approach would be broad mitonuclear concordance within asexual P. antipodarum such that asexuals are most similar to coexisting sexuals at both nuclear and mtDNA markers. This result would suggest recent origin of the asexuals from sympatric sexuals. Alternatively, a pattern where many asexuals harbour nuclear and/or mitochondrial genotypes that are different from coexisting sexuals will implicate a more complex scenario of asexual lineage origin, suggesting that at least some asexual lineages are widespread and of allopatric and/or nonrecent derivation from sexual P. antipodarum.

Based on this logic, we generated nuclear SNP and mitochondrial sequence data from P. antipodarum from 16 New Zealand lake populations to compare and contrast the population genetic structure of sexual and asexual P. antipodarum as reflected by both the mitochondrial and nuclear genomes. In particular, we address 1) the extent to which asexual lineages of P. antipodarum are of recent origin from coexisting sexuals, 2) how mitonuclear discordance is distributed in sexual vs. asexual individuals and 3) whether an especially common and widespread mitochondrial haplotype found primarily in asexual lineages is associated with common and widespread nuclear genotypes.

Materials and methods

Sample collections

We collected 574 snails in 10 New Zealand lakes in January 2009 and 6 New Zealand lakes in January 2010 (Table 1, Fig. 1) by pushing kick nets through vegetation and washing snails off rocks. All samples were collected in the shallow (0–1 m) littoral area of the lakes with the exception of Lake Alexandrina, where the mid-water (1–3 m) and deep-water (4–6 m) habitats were also sampled using snorkelling equipment. Shallow and deep habitats in this well-studied lake are known to differ in the relative frequencies of sexual and asexual Potamopyrgus antipodarum, with sexual diploids most common in the shallow and asexual triploids dominating the deep; the mid-water habitat harbours intermediate proportions of sexuals and asexuals (Fox et al. 1996). Including samples from the deeper habitats

© 2013 John Wiley & Sons Ltd
Table 1 Characteristics of the 16 sampled New Zealand lakes. Sample sizes (N) include only individuals for which we obtained a single-nucleotide polymorphism (SNP) genotype with no more than 11 missing loci (92.7% of sampled individuals); see Methods section ‘SNP marker development and genotyping’ for details.

<table>
<thead>
<tr>
<th>Population (acronym)</th>
<th>Geographic coordinates (Latitude, Longitude)</th>
<th>Year collected</th>
<th>N</th>
<th>N diploids</th>
<th>Percentage 2x</th>
<th>Percentage 2x males</th>
<th>N unique genotypes among diploids</th>
<th>Percentage 3x</th>
<th>Percentage &gt;3x</th>
<th>Percentage unknown ploidy</th>
<th>N unique lineages among polyploids</th>
<th>Haplotypes (percentage among individuals)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexandrina (AX)</td>
<td>−43.937784, 170.455286</td>
<td>2009</td>
<td>122</td>
<td>50</td>
<td>40.98</td>
<td>13.11</td>
<td>49</td>
<td>59.02</td>
<td>0.00</td>
<td>0.00</td>
<td>40</td>
<td>10A (15.6); 10A+1(B) (11.1); 1A (45.6); 25A (8.9); 30A (16.7); 30A+1(B) (1.1); 30C (4.4); 30E (5.6); 50 (1.1)</td>
</tr>
<tr>
<td>Clearwater (CW)</td>
<td>−43.6096176, 171.046179</td>
<td>2009</td>
<td>31</td>
<td>22</td>
<td>70.97</td>
<td>25.81</td>
<td>22</td>
<td>22.58</td>
<td>9.68</td>
<td>0.00</td>
<td>6</td>
<td>1A (21.4); 30A (75.0); 30A+1(A) (3.6)</td>
</tr>
<tr>
<td>Grasmere (GR)</td>
<td>−43.061572, 171.774601</td>
<td>2009</td>
<td>28</td>
<td>7</td>
<td>25.00</td>
<td>14.29</td>
<td>7</td>
<td>64.29</td>
<td>7.14</td>
<td>0.00</td>
<td>19</td>
<td>1A (73.1); 30A (26.9)</td>
</tr>
<tr>
<td>Gunn (GU)</td>
<td>−44.875886, 168.090282</td>
<td>2010</td>
<td>32</td>
<td>0</td>
<td>0.00</td>
<td>NA</td>
<td>NA</td>
<td>71.88</td>
<td>28.13</td>
<td>0.00</td>
<td>21</td>
<td>10A (14.3); 10A+1(A) (4.8); 10A+2(A) (9.5); 1A (23.8); 1A+1(A) (4.8); 20A (33.3); 37A (4.8); 68 (4.8)</td>
</tr>
<tr>
<td>Haupiri (HR)</td>
<td>−42.569563, 171.688418</td>
<td>2010</td>
<td>37</td>
<td>3</td>
<td>8.11</td>
<td>8.11</td>
<td>3</td>
<td>83.78</td>
<td>8.11</td>
<td>0.00</td>
<td>22</td>
<td>1A (88.0); 27A (12.0)</td>
</tr>
<tr>
<td>Heron (HE)</td>
<td>−43.479353, 171.172282</td>
<td>2009</td>
<td>38</td>
<td>30</td>
<td>78.95</td>
<td>23.68</td>
<td>30</td>
<td>15.79</td>
<td>5.26</td>
<td>0.00</td>
<td>7</td>
<td>1A (24.3); 30A (67.6); 52 (8.1)</td>
</tr>
<tr>
<td>Ianthe (IA)</td>
<td>−43.053442, 170.624366</td>
<td>2010</td>
<td>31</td>
<td>29</td>
<td>93.55</td>
<td>41.94</td>
<td>28</td>
<td>6.45</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
<td>1A (3.2); 37A (3.2); 41A/B + 1(A) (6.5); 41A/B + 2(A) (9.7); 41A/B + 2(B) (12.9); 41A/B + 3(B) (3.2); 41A/B + 4(A) (3.2); 41B (54.8); 41B+1(A) (3.2)</td>
</tr>
<tr>
<td>Mapourika (MP)</td>
<td>−43.31801, 170.204244</td>
<td>2010</td>
<td>33</td>
<td>7</td>
<td>21.21</td>
<td>3.03</td>
<td>7</td>
<td>24.24</td>
<td>0.00</td>
<td>54.55</td>
<td>5</td>
<td>1A (34.5); 1A+1(B) (6.9); 27A (37.9); 30D (17.2); 30D+1(A) (3.4)</td>
</tr>
<tr>
<td>Okareka (OK)</td>
<td>−38.171413, 176.361585</td>
<td>2009</td>
<td>13</td>
<td>5</td>
<td>38.46</td>
<td>0.00</td>
<td>4</td>
<td>0.00</td>
<td>0.00</td>
<td>61.54</td>
<td>0</td>
<td>37A (20.0); 61 (10.0); 62/64 + 1(A) (10.0); 68 (60.0)</td>
</tr>
<tr>
<td>Poerua (PR)</td>
<td>−42.704813, 171.495172</td>
<td>2009</td>
<td>16</td>
<td>1</td>
<td>6.25</td>
<td>0.00</td>
<td>1</td>
<td>56.25</td>
<td>37.50</td>
<td>0.00</td>
<td>8</td>
<td>1A (100.0)</td>
</tr>
<tr>
<td>Rotoiti (RT)</td>
<td>−41.820374, 172.837135</td>
<td>2009</td>
<td>18</td>
<td>0</td>
<td>0.00</td>
<td>NA</td>
<td>NA</td>
<td>27.78</td>
<td>72.22</td>
<td>0.00</td>
<td>12</td>
<td>1A (91.7); 37B (8.3)</td>
</tr>
<tr>
<td>Selfe (SE)</td>
<td>−43.240251, 171.519869</td>
<td>2009</td>
<td>33</td>
<td>16</td>
<td>48.48</td>
<td>24.24</td>
<td>16</td>
<td>51.52</td>
<td>0.00</td>
<td>0.00</td>
<td>13</td>
<td>1A (44.8); 27A (37.9); 27B (3.4); 9A (13.8)</td>
</tr>
<tr>
<td>South Mavora (MV)</td>
<td>−45.302545, 168.173651</td>
<td>2010</td>
<td>35</td>
<td>0</td>
<td>0.00</td>
<td>NA</td>
<td>NA</td>
<td>77.14</td>
<td>22.86</td>
<td>0.00</td>
<td>20</td>
<td>10A+1(A) (20.0); 1A (35.0); 1A+1(O) (5.0); 20A (25.0); 20A+1(A) (10.0); 20A+2(A) (5.0)</td>
</tr>
</tbody>
</table>
in Lake Alexandrina is unlikely to bias the across-lake results, as there are no significant differences in genetic similarity between sexual and asexual snails sampled from different habitats in this lake (D. Paczesniak, unpublished data). All snails were transported to the University of Iowa (2009) or the Swiss Federal Institute of Aquatic Science and Technology (Eawag) (2010) and housed in aquaria until further analysis (up to 4 weeks, Iowa; up to 8 weeks, Eawag). Individual snails were dissected under a dissecting microscope and sexed (males were distinguished from females by a presence of a penis). The head was split in two halves and either snap-frozen in separate tubes for DNA extraction and flow cytometry (Iowa) or used immediately for flow cytometry analysis (Eawag).

Flow cytometry

The intensity of fluorescence of cell nuclei extracted from snail heads and stained with DAPI (4',6-diamidino-2-phenylindole) relative to a size standard was used to assign ploidy level and thus reproductive mode (2× individuals are sexual, 3× and ≥3× individuals are asexual; Wallace 1992; Neiman et al. 2011, 2012). The only potential exceptions to the diploid-sexual association were made for diploids from lakes Okareka and Te Anau. We treated these snails differently because no diploid males have ever been collected from these lakes, suggesting that diploid females from Okareka and Te Anau may be in fact asexual (Neiman et al. 2011). Because the confirmation of asexual status of the diploids in these two lakes requires additional study, we excluded Okareka and Te Anau diploids from the sexual group in the two analyses in which we directly compared sexual and asexual snails (testing the influence of reproductive mode on the assignment of individuals to their population of origin by the clustering analysis and testing the association of 1A haplotype with asexuality). Samples collected in 2009 were analysed at the University of Iowa and those collected in 2010 at Eawag. The detailed laboratory protocols, methods of data analysis and methods used to combine the data obtained in the two different laboratories are presented in Neiman et al. (2011). Altogether, we were able to successfully assign ploidy in 504 of the 532 snails (94.7%) for which we obtained a SNP genotype (see below).

DNA extraction

We extracted DNA from snail head tissue with the Qiagen DNeasy Plant kit (Qiagen, Valencia, CA, USA) following the manufacturer’s protocol with one modification: we used nanopure water rather than the elution buffer for the final dilution.
SNP marker development and genotyping

For the purpose of this study, we used the *P. antipodarum* transcriptome (Wilton *et al.* 2013) to develop a set of nuclear SNP markers. We used two different strategies to isolate two different sets of contigs from the transcriptome for candidate SNP detection and selection:

1. We used contigs that returned a top hit to molluscs in a blastx search (Altschul *et al.* 1997) in GenBank. This strategy ensured that we did not use SNPs within sequences generated from microorganisms present on or in the *P. antipodarum* used for the transcriptome. Next, we used blastx to further exclude contigs with hits to gene families or gene isoforms in molluscs, which reduces the possibility of including members of multigene families.

2. We used contigs that matched single-gene families identified by Holton & Pisani (2010), who used metazoan genome data and identified single-gene families suitable for phylogeny reconstruction.

Within these two sets of contigs, we then identified variable nucleotide sites that represented candidate SNP sites. Next, we used blastx to assign reading frame and thus determine whether each putative SNP comprised a synonymous or nonsynonymous change. Because we wanted to focus on neutral (or nearly neutral) sites, we selected only synonymous SNPs. In the case of multiple SNPs per contig that satisfied our criteria, the target SNP was chosen randomly.

Sequences of 100 candidate SNPs along with 100–500 bp of the flanking sequence on either side were sent to the Genomic Technologies Facility at Iowa State University for design of the Mass Array® iPLEX® SNP genotyping assay and for the SNP genotyping itself. Out of the genotyping multiplex assays designed *in silico*, we chose the assay with the highest number of loci (32). Of the 32 genotyped loci, 23 were polymorphic and scorable in >70% of individuals genotyped; only these loci were used for further analysis. Of the 574 genotyped individuals, we excluded from further analysis 42 individuals with missing data at 12 or more loci, leaving our final data set with 532 individuals. For individuals of all ploidy levels, the genotypes at each locus were scored as homozygous when only one allele was present and as heterozygous when two alleles were present. It was not possible to reliably assign the identity of additional alleles in polyploid individuals (e.g. to distinguish between genotypes AAB and ABB).

In order to identify members of the same asexual lineage, we then used these multilocus SNP genotypes and GenoType software (Meirmans & Van Tienderen 2004) to assign genotypic identity to polyploid individuals. Missing data were not considered, and the threshold for assigning individuals as members of different lineages was set to zero, meaning that individuals with pairwise distances between genotypes larger than zero were scored as members of a different lineage. If individuals with identical SNP genotypes had different mitochondrial haplotypes or were of different ploidy level, they were manually assigned as members of a unique lineage. To account for the possibility that there may be rare diploid asexual individuals, we also identified identical SNP genotypes among diploid individuals in our samples using GenoType software with the same settings as described above.

Mitochondrial sequencing

A 718-bp portion of the mitochondrial cytochrome b gene had already been sequenced in 243 of the 532 *P. antipodarum* for which we had generated SNP genotypes (Neiman *et al.* 2011). We followed protocols used in Neiman *et al.* (2011) to sequence the same 718-bp segment in the additional 289 *P. antipodarum* included in the present study.

The newly acquired cytochrome b sequences were edited and aligned in BioEdit (Hall 1999). We then trimmed the sequences, where possible, to the same 718-bp fragment of the gene used in Neiman *et al.* (2011). Next, each new sequence was used as a query in a blastn search against NCBI's nucleotide database to determine whether the sequence was identical to one of the 45 different haplotypes identified in Neiman *et al.* (2011) (GenBank Accession nos JF518834-JF518877). If so, the haplotype was assigned the same name as the matching GenBank haplotype. If not, the haplotype was given the name of the closest match on GenBank plus the number of substitutions separating it from that haplotype.

After assigning the haplotype names, we trimmed the sequences to the 634 bases that were reliably amplified in all individuals and then used this data set in all subsequent analyses. To visualize the genetic structure of mitochondrial sequences, we constructed a maximum-parsimony haplotype network at the 95% connection limit using TCS version 1.21 (Clement *et al.* 2000). Next, we used BIC value comparison to select the model of sequence evolution that best fit the data. We then ran 1000 replicates of neighbour-joining and maximum-likelihood bootstrap procedures under the best-fitting model (the Tamura 3-parameter model: rate variation among sites was modelled using gamma distribution, shape parameter = 0.5066) to evaluate the statistical support of the resulting tree topologies. Both model...
selection and tree construction were performed using MEGA 5 (Tamura et al. 2011).

Genetic structure: mitochondrial and nuclear data

We first confined our analyses to the 414 unique genotype/ploidy/mitochondrial haplotype combinations in order to avoid biases arising from the treatment of genetically identical individuals from the same lake (i.e. members of the same asexual lineage) as independent data points. Next, we used calculations of pairwise \( F_{ST} \) values for the five populations for which we had a sufficient sample of sexual individuals to estimate \( F \) statistics \((N > 16\) individuals; Alexandrina, Clearwater, Heron, Ianthe and Selfe) to evaluate the ability of our SNP markers to detect genetic structure and to quantify differentiation between sexual populations. We also calculated the unbiased estimate of global \( F_{ST} \) \((0\), Weir & Cockerham 1984\) and bootstrap 95\% confidence limits using Fstat version 2.9.3 (Goudet 1995).

We used discriminant analysis of principal components (DAPC) (Jombart et al. 2010), as implemented in the R package adegenet (Jombart 2008; R Development Core Team 2009), to cluster genetically similar individuals according to their multilocus SNP genotypes. This method first transforms the data using principal components analysis, which ensures that the variables are not correlated and that the number of variables is smaller than the number of individuals. Then, discriminant analysis partitions the variance into among- and within-group components, maximizing separation between groups. Because DAPC does not assume any underlying population genetic model, it is applicable to a mixed sample of sexual and asexual individuals (Jombart et al. 2010).

We performed the DAPC analysis on the nuclear SNP data set using two sets of priors: 1) population of origin (16 lakes) and 2) mitochondrial haplotype group (seven groups, Figs 1 and 2). We defined these groups as those sets of haplotypes separated from one another by at least five mutational steps (>1\% pairwise divergence). Because the largest of these groups contained two distinct sets of haplotypes that were clearly differentiated on the basis of high relative frequency of sexuals (P3, Fig. 2) vs. asexuals (P5, Fig. 2), we treated these subgroups as separate groups for this analysis. The inclusion of these separate P3 and P5 groups allowed us to evaluate whether differences in nuclear population genetic structure were evident in haplotype groups that tend to differ in reproductive mode. Because group P5 also included a haplotype (1A) that is overwhelmingly common in South Island asexual \( P. \) antipodarum, this analysis also allowed us to address whether there was a specific nuclear population genetic structure associated with this haplotype. To compare the population genetic structure revealed by nuclear and mitochondrial data, we also performed the DAPC analysis on the mitochondrial haplotype data using population of origin as prior.

We performed these three DAPC analyses with these two types of priors because comparison of their ability to explain population genetic structure in sexuals vs. asexuals can provide important information on the origin and diversity of asexual lineages. First, the population of origin represents both the present mating pool for sexuals and the present selection arena for asexuals. These processes are expected to generate population genetic structure unless countered by migration (asexuals) and/or gene flow (sexuals and possibly asexuals). Our analysis of among-lake \( F_{ST} \) values in sexual \( P. \) antipodarum revealed substantial differentiation and thus low gene flow among sexual populations and also indicates that population of origin confers meaningful genetic information for sexuals (Table 2, also see Dybdahl & Lively 1996). The implications of this \( F_{ST} \) information are that we can use population of origin priors for the nuclear data to address whether population of origin also confers meaningful genetic information for asexual \( P. \) antipodarum. Given the evidence from our \( F_{ST} \) analyses that sexual \( P. \) antipodarum populations experience little gene flow (Table 2), we predict that if asexual \( P. \) antipodarum are of recent origin from sympatric sexuals, the ability of the population of origin prior to explain asexual nuclear population genetic structure should be similar to or greater than the ability of this prior to explain structure in sexual \( P. \) antipodarum.

In contrast to population of origin, mitochondrial haplotype group membership reflects historical events (e.g. Pleistocene glaciation, Trewick et al. 2000; Neiman & Lively 2004) that occurred thousands to millions of years ago. This means that individuals that share a mitochondrial haplotype group – regardless of current population of origin – are likely to also share at least some of their evolutionary history. Inclusion of both population of origin and mitochondrial haplotype group priors thus gives us information on the evolutionary histories of sexual and asexual \( P. \) antipodarum

<table>
<thead>
<tr>
<th></th>
<th>Alexandrina</th>
<th>Clearwater</th>
<th>Heron</th>
<th>Ianthe</th>
<th>Selfe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexandrina</td>
<td>0.2231</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clearwater</td>
<td>0.2373</td>
<td>0.1076</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heron</td>
<td>0.4061</td>
<td>0.4439</td>
<td>0.327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ianthe</td>
<td>0.2313</td>
<td>0.2611</td>
<td>0.2476</td>
<td>0.4103</td>
<td></td>
</tr>
<tr>
<td>Selfe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
across a wide timescale. We can also use comparisons between outcomes of the DAPC analyses with population of origin prior between the nuclear and mitochondrial data sets to provide information on the extent to which there is mitochondrial vs. nuclear population genetic structure across populations and in diploids (all sexual with possible exception of snails from Te Anau and Okareka) vs. asexual polyploids.

We can evaluate how well our data fit these various predictions by comparing our DAPC analyses results to a null prediction: if there was no structure in the data that can be explained by our priors, then each individual should have the same probability of being placed in any of the posterior groups (1/16 for the population prior, 1/7 for the haplotype group prior). Based on this logic, we compared the mean value of the maximum probability that an individual belonged to any of the posterior groups assigned by DAPC analysis to the null values of 1/16 (for analyses using population prior) and 1/7 (haplotype group prior) using 95% confidence intervals (obtained from 10 000 bootstrap replicates). We analysed means of all individuals together and also conducted separate analyses for each ploidy level. We performed the DAPC analyses with the nuclear SNP data using the population of origin and haplotype group priors and with the mitochondrial haplotype data using the population of origin prior.

Finally, we used information from the DAPC analysis of the nuclear SNP data with the predefined population of origin prior to construct a contingency table evaluating whether and how often population of origin correctly predicted cluster membership. We then used logit models to test whether a significantly higher proportion of sexual vs. asexual individuals was correctly reassigned to their prior groups (population of origin). In a logit model, the binomial dependent variable (correctly vs. incorrectly reassigned) can be explained with categorical independent variables; here, population and reproductive mode. We used this logit analysis for the nine populations where both diploid sexuals and polyploid asexuals were present to evaluate whether the presence of 1A was independent of the population (lake) and/or reproductive mode. The analyses were performed using IBM® SPSS Statistics version 19 and Microsoft® Excel 2010.

A role for selection? All else being equal, genetic diversity in the nuclear and mitochondrial genomes is expected to be proportional to the effective population size experienced during the recent evolutionary history of the genome (Gompert et al. 2008). The implications are that insight into why a particular haplotype is over-represented can come from comparisons of the levels of nuclear and mitochondrial genetic diversity in populations with varying proportions of the haplotype in question. As any mitochondrial haplotype becomes relatively common in a population, mitochondrial genetic diversity will automatically become relatively low. Nuclear genetic diversity, however, will only decrease as the frequency of a particular mtDNA haplotype becomes common if similar evolutionary forces were affecting both genomes – for example, a recent population bottleneck. Conversely, retention of high nuclear diversity in populations where the haplotype is common (and mitochondrial diversity is low) is predicted under positive selection favouring a mitochondrial genome and/or cytoplasm bearing a particular mitochondrial haplotype (Gompert et al. 2008). By this logic, we used estimates of nuclear diversity in the study populations to determine whether relatively high frequency of the 1A haplotype (and thus relatively low mitochondrial diversity) is associated with reduction in genetic diversity in nuclear genomes.

As a measure of nuclear genetic diversity we used the mean of pairwise comparisons between individual genotypes with the ‘band-sharing’ pairwise genetic identity measure that was developed by Tomituk et al. (2009) to compare intraspecific genetic differentiation. This identity measure can be used for samples that vary in ploidy level. The mean of pairwise genetic identities between individual genotypes from the same lake provides an estimate of the genetic diversity of the population, with values closer to one indicating less diverse populations (individuals are more similar to each other) and values closer to 0 indicating more diverse populations (individuals are more dissimilar to each other).
We calculated these pairwise genetic identity measures with POPDIST version 1.2.4 (Tomituk et al. 2009).

We then used a linear regression model (as implemented within IBM® SPSS Statistics version 19) to examine the relationship between nuclear genetic diversity and the frequency of the 1A haplotype. This analysis allowed us to examine whether nuclear diversity was reduced in populations with relatively high frequency of the 1A haplotype (and thus relatively low mitochondrial diversity).

Results

Genetic structure: mitochondrial and nuclear data

We identified 52 unique mitochondrial haplotypes. The haplotype network at 95% connectivity separated these haplotypes into three networks (Fig. 2) that correspond to the major groups previously described in Potamopyrgus antipodarum (Neiman & Lively 2004; Neiman et al. 2011). Similar to these previous studies, we found that a single recently derived mitochondrial haplotype (1A) was very common relative to other haplotypes, found in 36% of all individuals and in 12 of 16 lakes (Table 1).

The same major divisions identified by the haplotype network procedure were also visible in the neighbour-joining tree (Fig. 1). The maximum-likelihood method yielded the same tree topology with very similar bootstrap values (tree not shown).

The five samples for which we had the largest samples of diploid sexual genotypes (N > 16) show substantial population differentiation (Table 2), with a mean $F_{ST}$ of 0.29 and standard deviation of 0.11. The global $F_{ST}$ estimate $\theta = 0.298$ (95% confidence limits: lower CL = 0.203, upper CL = 0.390) is somewhat higher than the $\theta = 0.128$ (95% confidence limits: lower CL = 0.031, upper CL = 0.225) estimated with allozyme data by Dybdahl & Lively (1996), which might be explained by the larger and more geographically widespread set of lakes we sampled.

To visualize the phylogeographic structure of the nuclear data, we plotted the result of the DAPC analysis with population of origin as prior (Fig. 3). While there are no clear separations into nonoverlapping groups, there is clustering of samples from the same geographic region. For example, the separation of samples is mostly along the first axis (LD1), which differentiates the east coast of the South Island from the other regions.

We then compared the patterns of divergence revealed by the DAPC analyses using population of origin vs. mitochondrial haplotype group as priors (Fig. 4). Lakes for which the two analysis outcomes are similar represent populations that have been isolated long enough that the lake of origin and mitochondrial haplogroup priors are similarly informative regarding DAPC cluster assignment (e.g. Te Anau, Ianthe). For South Island lakes, the DAPC analyses with population of origin and mitochondrial prior reveal a west coast/east coast division as well as fine-scale structure among lakes.

Next, we compared DAPC analysis outcomes for sexual diploids vs. asexual polyploids (Fig. 4). While the asexuals are often quite similar to coexisting sexuals (e.g. Mapourika, Clearwater), differentiation between coexisting sexuals and asexuals is apparent in several lakes. This differentiation is especially notable for the analysis with the population of origin prior for samples from Alexandrina, Selfe and Grasmere, where many asexual individuals have nuclear genotypes that are not found in sympatric sexuals (Fig. 4, top panel).

We then used comparisons of the 95% confidence intervals around the mean maximum probability of membership in a DAPC-assigned cluster relative to the null assumption of membership probability if priors convey no useful information (equal across all clusters) to determine whether either or both of the population of origin and mitochondrial haplotype group priors provided useful information regarding genetic structure (Fig. 5).

While both priors explained significantly more structure than by chance alone (Fig. 5A), the population of origin prior was more than twice better at predicting cluster membership (13.1x relative to null) than the mitochondrial haplotype prior (5.2x relative to null). The population of origin prior also significantly exceeded null expectations for the mitochondrial haplotype data, but to a lesser extent (9.5x relative to null) than for the nuclear data.

The same analysis performed for the different ploidy levels showed significantly lower values of mean maximum cluster membership for both triploid and >3x individuals vs. diploid individuals with the population of origin prior and for both the nuclear and mitochondrial data, but we found no significant differences between triploid and >3x individuals for any of the comparisons (Fig. 5B). These results suggest that there is more population genetic structure for diploid P. antipodarum, which are all or nearly all sexual, than for polyploid asexual P. antipodarum. There were no significant differences between any of the pairwise comparisons of ploidy levels in the analysis using nuclear data with mitochondrial group prior, meaning that the extent of nuclear structure that can be explained by mitochondrial haplotype group does not differ between the different ploidy levels.

We then used logit models to ask whether assignment of individuals to their population of origin
differed between sexual and asexual snails. We first compared the fit of a model with main effects of lake of origin and reproductive mode to a full model including both main effects and the interaction term. Because the full model did not fit the data significantly better than the reduced model ($\chi^2 = 15.38, P = 0.052, df = 8$), suggesting that the amount of variation explained by the interaction term was not very high, we used the reduced model (main effects only) to evaluate the relative strength of reproductive mode vs. lake of origin as explanations for the correct assignment of *P. antipodarum* to their prior group (population of origin). 

![Diagram](image_url)
Fig. 2 Maximum-parsimony network of mitochondrial cytochrome b haplotypes. The TCS analysis at the 95% connection limit identified three networks. The area of each pie chart is approximately proportional to the frequency of the haplotype, and the proportion of each of 4 types of shading represents the proportion of individuals with that haplotype represented by each of three possible ploidy levels as well as individuals with unknown ploidy. Each line segment represents a single nucleotide substitution, and the nodes represent unsampled haplotypes. Haplotypes encircled with dashed lines represent seven haplotype groups (P1-P7) used as priors in DAPC analysis (see Methods for details).
found that both lake ($\chi^2 = 20.19, P = 0.010, \text{df} = 8$) and especially reproductive mode ($\chi^2 = 20.19, P = 7 \times 10^{-6}, \text{df} = 1$) significantly explained the proportions of correctly reassigned individuals. In particular, a higher overall proportion of sexuals (82.2%) than asexuals (61.5%) were correctly assigned to their population of origin by the DAPC analysis (Fig. 6). This result provides another line of evidence for a higher level of population genetic structure for diploid sexual vs. polyploid asexual *P. antipodarum*.

Tests for a) association of the common 1A haplotype with asexuality and b) a mitochondrial selective sweep

We first compared the ability of a logit model with main effects of lake of origin and reproductive mode and a full model including both main effects and the effect of the interaction term to explain the distribution of the 1A haplotype within and across lakes, and found no significant differences between the models ($\chi^2 = 9.81, P = 0.279, \text{df} = 8$). Thus, we used reduced models to evaluate whether there were significant main effects of lake of origin and/or reproductive mode (sexual vs. asexual) on the frequency of *P. antipodarum* bearing haplotype 1A. We found that both lake ($\chi^2 = 16.90, P = 0.031, \text{df} = 8$) and especially reproductive mode ($\chi^2 = 204.66, P = 2 \times 10^{-46}, \text{df} = 1$) influenced 1A frequency, with 1A much more likely to be found in asexual individuals.

The results of our regression analysis comparing across-population haplotype 1A frequency and nuclear genetic diversity showed that increased frequency of haplotype 1A was not associated with reduced nuclear genetic diversity ($R^2 = 0.0002, F_{1,14} = 0.003, P = 0.958$). This result suggests that genetic drift, which is expected to reduce diversity in both mitochondrial and nuclear genomes, is an unlikely cause of the commonness of haplotype 1A in asexuals, and suggests instead that selection on the mitochondrial genome and/or other cytoplasmic elements might be driving this pattern.

Discussion

Our nuclear SNP data revealed that asexual *Potamopyrgus antipodarum* harbour high genetic diversity and are often – but not always – closely related to coexisting sexual *P. antipodarum*. We also found that population genetic structure is strongly associated with lake of origin for both sexual and asexual *P. antipodarum*. These results are in part consistent with an earlier allozyme-based study indicating that asexual *P. antipodarum* are generally the product of multiple recent independent transitions to asexuality from sympatric sexual *P. antipodarum* (Dybdahl & Lively 1995), but depart from the
findings of this previous study in revealing evidence for allopatric and/or nonrecent origins for at least some asexuals. Consistent with this possibility, we also detected considerably more population structure for sexual than asexual \textit{P. antipodarum} and found that asexual \textit{P. antipodarum} were more likely than sexuals to harbour nuclear genotypes that did not group with other individuals from their lake of origin in a clustering analysis. We also found that a very common, geographically widespread and recently derived cytochrome b haplotype (1A) was strongly associated with asexuality (also see Neiman & Lively 2004; Neiman \textit{et al.} 2011). Counter to predictions that the mitochondrial and nuclear genomes should show congruence within asexuals, asexual individuals harbouring haplotype 1A often (but not always) had nuclear genotypes that were similar to those found in coexisting sexuals, indicating strong mitonuclear discordance within asexuals. Finally, higher relative frequency of this haplotype was not associated with reduced nuclear genetic diversity, suggesting that genetic drift alone is an unlikely cause of the commonness of 1A and pointing instead to a role for selection. Altogether, these data indicate that there is widespread mitonuclear discordance in asexual New Zealand \textit{P. antipodarum}, and hint that at least some of this discordance may be linked to the spread of the cytoplasm containing haplotype 1A in asexuals.

Previous mtDNA-based studies of the phylogeography of \textit{P. antipodarum} indicate that major genetic differences occur on a north/south axis (Neiman & Lively 2004; Neiman \textit{et al.} 2011). This pattern was also evident in our mitochondrial data (Fig. 1). In contrast, the main phylogeographic pattern revealed by our nuclear SNP data was genetic differentiation between the east part of the South island and other regions, though a north–south division was still apparent. Altogether, these patterns are consistent with the major phylogeographic barriers described in other New Zealand taxa (Wardle 1963; McGlone 1985; Heads 1998; Trewick \textit{et al.} 2000; Trewick & Wallis 2001): the division between North and South Islands and between the eastern and western drainages of the Southern Alps on the South Island.
Structure of asexual *P. antipodarum* populations

The evolutionary consequences of migration are expected to differ for sexual and asexual populations because only the sexuals should experience introgression of migrant alleles. Migrant asexual lineages should instead contribute whole genotypes to the recipient population, which might be expected to leave a stronger signal in population genetic structure analyses than introgression of individual migrant alleles. Such a difference in the detectability of sexual vs. asexual migrants could help to explain why the proportion of asexual polyploid *P. antipodarum* correctly assigned to their lake of origin by the clustering analysis was significantly lower than for diploid sexuals.

Regardless of potential detectability differences, our data suggest that at least some asexual *P. antipodarum* possess nuclear genotypes and/or mitochondrial haplotypes that are not of recent origin from coexisting sexuals. One possible explanation for this result is occasional ecological perturbations of sexual populations, which then recover to new genetic equilibria. It is plausible that it would take some time for asexual populations to reflect the new equilibrium allele frequencies of the coexisting sexual population, which could itself complicate cluster assignment of older asexual lineages representing preperturbation genetic equilibria. By this logic, we would predict more divergence between sexual and asexual lineages in environments where ecological perturbations are more likely. Because this hypothesis also posits that asexual individuals with genotypes that do not cluster with other individuals from their source populations represent relictual asexual lineages derived from relatively ancient (and perhaps extinct) sexual populations, we would also predict that these asexuals should show significantly more genetic divergence from any sexual *P. antipodarum* than asexual *P. antipodarum* that do cluster with other individuals from their lake of origin.

A different but nonmutually exclusive potential explanation for the existence of many asexual *P. antipodarum* with mitochondrial haplotypes and nuclear genotypes that do not readily group with other members of their source populations is that some of the asexual lineages are migrants derived from allopatric sexual populations. In contrast to the ‘perturbation’ hypothesis detailed above, this ‘migrant’ hypothesis predicts no consistent differences in genetic divergence from extant sexuals between ‘migrant’ and ‘nonmigrant’ asexuals. The implications are that one could discriminate between the relative importance of these two hypotheses by using a coalescent-based analysis to compare the genetic divergence from sexual *P. antipodarum* of asexual *P. antipodarum* that cluster with their source population vs. asexual *P. antipodarum* that do not. Rigorous execution of this analysis would require both sexual individuals sampled from many more populations than those included here as well as nuclear markers (e.g. microsatellites, sequence-based markers or a large number of SNP markers) that
sexual *P. antipodarum* carrying 1A are more likely to give rise to new asexual lineages or where the asexual lineages carrying 1A have higher fitness than sexual *P. antipodarum* and/or other asexual lineages.

Our results highlight the importance of using data from both mitochondrial and nuclear genomes and from a large and geographically representative sample to characterize population genetic structure. In our case these data enabled important new insights into the evolutionary history of sexual and asexual *P. antipodarum* that were not revealed by earlier mtDNA (Neiman & Lively 2004; Neiman et al. 2011) or allozyme-based studies (Dybdahl & Lively 1995). These data also illuminate the existence of intraspecific mitonuclear discordance in asexual lineages of *P. antipodarum*, which is surprising in light of the expectation that these genomes (unlike in sexuals) should be coinherited. This result begs the question of the source of this variation, especially given the strong association between asexuality and a particular mitochondrial haplotype that our data suggest may have spread via selection. Evidence for widespread mitonuclear discordance in asexual organisms also sets the stage for the possibility of occasional violations of assumed strict asexuality (Schurko et al. 2009; also see Neiman et al. 2011).

**Acknowledgements**

We thank T. Holton and D. Pisani for sharing data files of the single-copy gene families; P. Wilton for assistance in developing the SNP loci; M. Möst for suggesting the DAPC analysis; P. Meirmans and T. Jombart for discussion of data analyses; J. Antonovics, L. Rice, J. Sharbrough, C. Tucci, M. Möst, A. Davison and three anonymous reviewers for comments on the previous version on the manuscript; and C.M. Lively, K. Klappert, K. King and K. Theisen for help with sample collections. We also acknowledge M. Wilkening and W. Wu for their assistance in developing our SNP assays and G. Hehman and C. Toll for help with cytochrome b sequencing. This work was supported by the Swiss National Science Foundation (JJ, grant #31003A_129961/1), the Carver Trust (MN), the Norwegian Research Council (MN, grant #196468/V40), the Iowa Center for Research by Undergraduates (KL) and the National Science Foundation (MN, MCB-1122176).

**References**


D.P., J.J. and M.N. designed research, analysed data and wrote the article. K.L. contributed to labwork and phylogenetic analyses. J.J. and M.N. funded research.

Data accessibility

DNA sequences: GenBank Accession nos: KF202473-KF202497. Note that in order to comply with GenBank requirements for sequence names we changed ‘+’ to ‘:’; ‘/’ to ‘.’ and ‘(’ to ‘*’ before the letter in the submitted haplotypes. For example, haplotype ‘41A/B + 3(B)’ is called ‘41A:B:3*B’ in GenBank.

SNP genotype, ploidy level data, sequence alignments: Dryad doi:10.5061/dryad.j18pv