*Genet. Res., Camb.* (2003), **81**, *pp.* 193–204. *With 2 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006219 Printed in the United Kingdom


# The influence of demography, population structure and selection on molecular diversity in the selfing freshwater snail *Biomphalaria pfeifferi*

BERNARD ANGERS[1]†, NATHALIE CHARBONNEL[1], NICOLAS GALTIER[2] AND PHILIPPE JARNE[1]*
[1] *Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de la Recherche Scientifique, Montpellier, France*
[2] *Genomes, Populations, Interactions, Centre National de la Recherche Scientifique – Université Montpellier II, Montpellier, France*

(*Received 15 May 2002 and in revised form 2 December 2002*)

**Summary**

Several forces may affect the distribution of genetic diversity in natural populations when compared to what is expected in a random-mating, constant size population of neutral genes. One solution for unravelling their respective influence is to study several genes at once in order to better reflect the true genealogy. Here we reconstruct the evolutionary history of the freshwater snail *Biomphalaria pfeifferi* over its entire distribution, using eight African populations, and three congeneric species as outgroups. A phylogenetic analysis was conducted using amplified fragment length polymorphism markers, and sequences at eight nuclear non-coding loci and one mitochondrial gene were used to analyse population structure. The geographic distribution of variation suggests greater affinities within than among regions. The pattern of variability at both the nuclear and mitochondrial DNA (mtDNA) loci is consistent with a bottleneck, although population structure may also partly explain our results. Our results are also indicative of the role of selection, whether positive or purifying, in the mtDNA. This highlights the fact that the interfering influences of population structure, demography and selection on molecular variation are not easily distinguished.

## 1. Introduction

Inferring the evolutionary history of populations is one main goal of modern population genetics (see e.g. Hartl & Clark, 1997; Singh & Krimbas, 2000). From an empirical point of view, analyses are based on genetic markers such as microsatellite loci or gene sequences from both nuclear and extra-nuclear genomes (see review in Goldstein & Schlötterer, 1999; Avise, 2000). Data are then analysed using population genetic models, based for example on coalescent theory. The simplest situation modelled is a random-mating population with stable demography at genetic equilibrium between mutation and genetic drift. However mating is not always random, populations may expand or crash and/or be connected by gene flow, while genes may experience the influence of selection. These processes strongly affect the distribution of coalescence times, and thus of variability within and among individuals and populations. They must therefore be taken into account when interpreting population history (see e.g. Begun & Aquadro, 1991; Charlesworth *et al.*, 1993; Takahata *et al.*, 2001; Wilkins & Wakeley, 2002 and references therein).

Our study is concerned with the influence of these processes, namely demographic variation, population structure and selection, on genetic variability. (i) Demographic variation may be considered on various geographic and temporal scales (e.g. local variation within populations versus species expansion), and our concern is with large-scale variation. A species may expand its range and number of individuals (expansion), or be restricted to a smaller area and number of individuals (bottlenecks). Such long-term demographic effects have been extensively explored in genetic analyses of human populations (e.g. Harpending *et al.*, 1998; Hey & Harris, 1999; Takahata *et al.*, 2001), and are attracting increasing interest in phylogeographic studies (review in Hare, 2001). Strong bottlenecks cause phylogenies to have shorter external and longer internal branches. (ii) Population structure

* Corresponding author. CEFE-CNRS, 1919 route de Mende, 34293 Montpellier cedex 5, France. e-mail: jarne@cefe.cnrs-mop.fr
† Present address: Département des Sciences Biologiques, Université de Montréal, Montréal, Canada.

strongly affects genetic variability, as the balance between genetic drift, gene flow and rates of extinction/recolonization defines the effective size of populations at local and global scales (review in Pannell & Charlesworth, 2000). Numerous models of population structure can be thought of, and it is therefore hard to predict its influence on the distribution of variability, especially in terms of coalescent topology (e.g. Nordborg, 2001; Wilkins & Wakeley, 2002). (iii) Genetic variability is also strongly affected by selection, especially when recombination is limited (Begun & Aquadro, 1991; Charlesworth *et al*., 1993), and numerous studies show that genes are under the influence of selection. For example, the review of Weinreich & Rand (2000) strongly suggests that mitochondrial genes experience purifying selection and occasionally selective sweeps (but see Gerber *et al*., 2001). Selection may take several forms, and we are here concerned with purifying and positive selection.

The question is how to disentangle the influence of these three processes on genetic variation, since they might have similar consequences on the distribution of variability. For example, bottlenecks and selective sweeps similarly affect phylogenetic reconstruction if only a single linkage group is considered (see e.g. Galtier *et al*., 2000). Purifying and positive selection both tend to push mutations towards external branches of coalescent trees (Williamson & Orive, 2002). Non-genetic data are certainly useful, such as results from studies of population dynamics, or historical records of expansion/contraction. Much can also be learned from the analysis of sequences, using for example tests of neutrality or demographic variation (review in Nielsen, 2001). A phylogenetic analysis within and among species is also extremely useful, since it allows one to evaluate the number of mutations along branches and infer common ancestors (Slatkin & Maddison, 1989). Probably the most useful approach is to analyse several genes from the same individuals and populations in a single analysis. This has been performed in humans (references in Cann, 2001; Takahata *et al*., 2001) and model organisms such as *Drosophila* (e.g. Schlötterer *et al*., 1997; Nurminsky *et al*., 2001) or the house mouse (Boissinot & Boursot, 1997). For example, selective sweeps are unlikely to affect more than one locus, while all loci should be affected by demographic processes or population structure.

The study presented here aims at reconstructing the evolutionary history of the freshwater snail *Biomphalaria pfeifferi*, the main intermediate host of *Schistosoma mansoni*, the agent of intestinal bilharziasis, in Africa. The analysis is based on nine loci – seven nuclear microsatellites, the first internal transcribed spacer (ITS) of ribosomal DNA and one mitochondrial DNA (mtDNA) gene – sequenced in populations from eight countries sampled over the species' current distribution. The study also includes sequence analysis of several populations from closely related species. We used amplified fragment length polymorphism (AFLP; Vos *et al*., 1995) markers to build a phylogeny that is independent of the loci sequenced. The questions addressed are: (i) Is there any evidence for large temporal scale demographic variation (bottleneck or expansion)? (ii) Are all loci studied equivalent with regard to selection? The null hypothesis was that they are, and the competing hypothesis was that the mtDNA locus has been under selection (other loci are either non-coding DNA, or not translated). The influence of population structure on the interpretation of data is also considered.

## 2. Materials and methods

### (i) *Species studied and sampled sites*

*B. pfeifferi* is a simultaneous hermaphroditic freshwater snail, whose distribution covers most of Africa south of the Sahara, Madagascar, and part of the Middle East (Brown, 1994). It occupies both transient and patchily distributed habitats (see e.g. Brown, 1994), which imposes strong demographic variation on populations (Woolhouse, 1992). The mean selfing rate in natural populations of *B. pfeifferi* is of the order of 90% (Charbonnel *et al*., 2002*a*, *b* and references therein), which should strongly influence the distribution of genetic variability within and among populations (Charlesworth *et al*., 1993; Jarne, 1995). However, no differential effect of selfing on variability is expected here since there is limited variation in selfing rate among *B. pfeifferi* populations. Limited migration has been detected among populations at large geographic distances (Bandoni *et al*., 1990; Mimpfoundi & Greer, 1990; Charbonnel *et al*., 2002*a*).

Individuals from three geographic regions were analysed (Table 1; Fig. 1). The western group included one site from Senegal, Ivory Coast, Niger and Cameroon, while the northeastern group included sites from Ethiopia and Oman and the southeastern group included one site from Zimbabwe and Madagascar (two sites from Madagascar for the AFLP analysis). A previous analysis at 10 microsatellite loci (Charbonnel *et al*., 2002*a*; B. Angers, unpublished data) indicated either no, or little, variation in the populations studied. Based on this information, a single individual was sequenced in monomorphic populations (two for the AFLP analysis), while several individuals displaying different microsatellite alleles were sequenced in polymorphic populations (see below for more details on sequencing). No sequence variation, except for the number of repeats, was detected within populations. Individuals from three additional species of *Biomphalaria* were included to serve as outgroups in the

Table 1. *Origin of the populations studied*

| Species | Group | Country | Site | Sender/sampler |
|---|---|---|---|---|
| *B. pfeifferi* | Northwestern | Cameroon | Lake Ngoa | F. Njiokou |
| | | Ivory Coast | Akakronzipris | B. Delay |
| | | Niger | Niamey | B. Delay |
| | | Senegal | BM 1971[a] | D. Rollinson |
| | Northeastern | Ethiopia | BM 1488[a] | D. Rollinson |
| | | Oman | Wadi | J.-P. Pointier |
| | Southern | Madagascar | Irondro | Present authors |
| | | Zimbabwe | Harare | S. Chandiwana |
| *B. alexandrina* | – | Egypt | Gizea | J.-P. Pointier |
| *B. sudanica* | – | Burundi | BM 1091[a] | D. Rollinson |
| *B. glabrata* | – | Venezuela | Lake Valencia | J.-P. Pointier |

The species, group (for *B. pfeifferi*), country and site of sampling are indicated for each population, as well as the name of the person who sampled or sent snails. See also Fig. 1.
[a] Refers to laboratory strains.

phylogenetic analysis: *B. alexandrina* (one site from Egypt), *B. sudanica* (one site from Burundi) and *B. glabrata* (three sites from Brazil, the Guadeloupe Island and Venezuela). A single individual was considered per site. Total DNA was extracted from frozen or ethanol-preserved individuals using the QIAmp tissues kit (Qiagen) according to the manufacturer's protocols or using a rapid extraction protocol (Valsecchi, 1998).

(ii) *AFLP screening*

The AFLP technique (Vos *et al.*, 1995) was used to screen a large number of (potentially polymorphic) fragments throughout the genome of *B. pfeifferi*, in order to build a phylogeny of populations and species independently of the other DNA fragments studied. The largely dominant nature of AFLP markers is generally considered undesirable in population genetics analyses (see e.g. Lynch & Milligan, 1994). However, a large fraction of loci in highly selfing species such as *B. pfeifferi* are homozygous, and individuals exhibiting a given band can safely be considered homozygous for this band. AFLP patterns were generated following Vos *et al.* (1995). Briefly, DNA was double-digested with *Mse*I and *Eco*RI. Digestion was stopped by heating at 70 °C for 15 min, and restriction fragments were ligated to the appropriate adaptors (Vos *et al.*, 1995). Selective amplifications were performed using the polymerase chain reaction (PCR) with eight pairs of primer combinations: *Eco*RI–AGG or *Eco*RI–ACC was used in combination with either *Mse*I–CAC, *Mse*I–CAG, *Mse*I–CTC or *Mse*I–CTG. Primers were labelled using $\gamma^{33}$P-ATP. The PCR mix was first heated at 92 °C for 30 s, prior to 13 touch-down cycles of 15 s at 92 °C, 30 s at an initial temperature of 65 °C decreased by 0·7 °C per cycle and 90 s at 70 °C. This was followed by 24 additional cycles of 15 s at 92 °C, 30 s at 56 °C and 90 s at 70 °C. Only those bands that showed both consistent amplification, and occurred in at least two populations, were retained for the analysis. *B. sudanica* was used as an outgroup in this analysis.

(iii) *DNA sequences*

Nine DNA regions were sequenced in all individuals. This includes 386 base pairs (bp) of the 16S rDNA region of the mitochondrial genome, the first ITS of the nuclear rDNA (ITS1 located between the 18S and the 5.6S regions; 640 bp), and the flanking regions (ranging from 137 to 338 bp) of seven microsatellite loci (Table 2). For the microsatellite loci, primers were redesigned based on the sequences of Charbonnel *et al.* (2000) in order to amplify sufficiently large regions (Table 1). However, amplification failed in *B. glabrata* for loci Bpf4, Bpf5 and Bpf8. The repeated regions of the microsatellite loci are not considered here. DNA fragments were amplified using PCR. The 100 $\mu$l reaction mix contained 100 pmol of target DNA, 50 pmol of primers, 5 $\mu$l of 10 × reaction buffer, 15 mM $MgCl_2$; 200 $\mu$M of each dNTP and 1 U of *Taq* polymerase. Primer sequences are given in Table 2. The PCR products were purified using the Wizard PCR DNA Purification System (Promega). Sequencing was done on both DNA strands with the fmol DNA sequencing system (Promega) using $\gamma^{33}$P-ATP. Sequences can be found in GenBank under references AY126519–AY126598.

(iv) *Data analysis*

AFLP data were analysed considering each band as a presence/absence character. The evolutionary relationships among populations were inferred by phylogenetic tree construction using both Wagner (Kluge & Farris, 1969) and Dollo (Farris, 1977) parsimony methods. The Wagner method assumes that

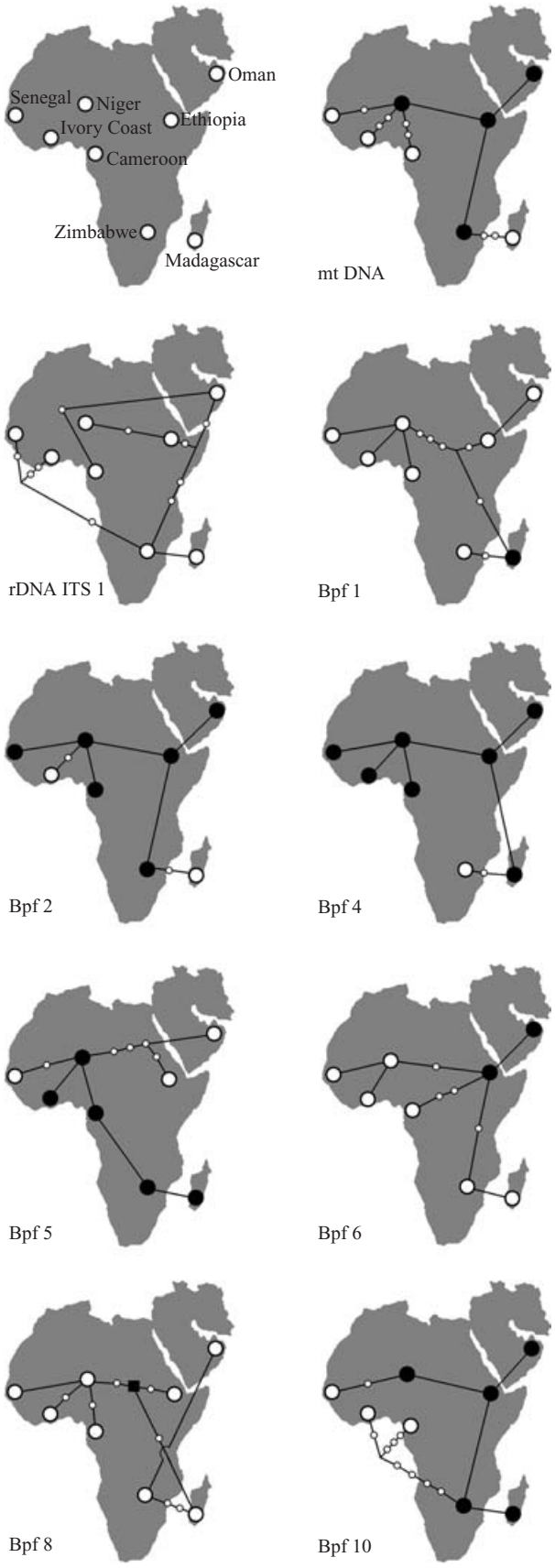


Fig. 1. Locations of the populations (large circles) studied across Africa, Madagascar and the Middle East (upper left map). See Table 1 for more details. Minimum-spanning networks among the haplotypes sequenced in *B. pfeifferi* are given on the following maps (one per locus). The small circles correspond to the number of substitutions (or indels) separating haplotypes (when the same haplotype occurs in more than one population, there are no 'small circles' on the lines connecting these populations). Filled large circles indicate the location of the MRCA which was always observed in the populations analysed, except for Bpf8 where MRCA is indicated by a square and ITS1 where MRCA was not inferred (see Materials and methods).

changes between two possible states have equal probabilities, while trait loss is more likely than trait gain in Dollo parsimony, which is more compatible with mutation at AFLP markers. The robustness of nodes was estimated by bootstrapping (1000 replicates). All phylogenetic analyses were performed using PHYLIP version 3.5c (Felsenstein, 1994). Trees were visualized using TREEVIEW version 1.6.1 (Page, 1996).

Sequences were aligned using CLUSTALW version 1.8 (Thompson *et al*., 1994) and optimized by visual inspection. Given the close relatedness among individuals, alignments were unambiguous, except for ITS1. Insertions and deletions of any length were treated as single characters. Several classical indices of diversity (Li, 1997, pp. 237–239) were estimated: the observed ($n_a$) and expected ($n_e$) number of haplotypes or alleles, the average number of nucleotide differences between pairs of sequences ($\Pi$), the nucleotide diversity ($\pi$) and the number of segregating sites ($K$). We also estimated $D_{p/as}$ and $D_{a/s}$ as the number of fixed sites between the ancestors of *B. pfeifferi* and *B. alexandrina–B. sudanica* sequences and in the *B. alexandrina–B. sudanica* group respectively. For the five loci for which it was possible to root the tree using *B. glabrata* as an outgroup (see the paragraph below on inference of networks), the numbers of sites separating the root from the ancestor of *B. pfeifferi* and from the ancestor of *B. alexandrina–B. sudanica* respectively was estimated. We also estimated the numbers of sites on the branches going to either *B. alexandrina* or *B. sudanica*. Estimating these parameters allowed us to produce various estimates of $\theta$ (the product of the effective population $N_e$ and the mutation rate $u$), which were used in neutrality tests (see below).

The analysis of the distribution of sequence diversity among populations and groups of populations (see above) was conducted using an analysis of molecular variance (AMOVA), with and without taking into account the number of nucleotide differences between haplotypes (Excoffier *et al*., 1992). The differentiation among groups of populations was quantified using both $\Phi_{st}$ and $F_{st}$, and their significance was assessed using a bootstrap method. This

https://doi.org/10.1017/S0016672303006219 Published online by Cambridge University Press

Table 2. *Loci studied, primer sequences and PCR conditions*

| Locus | Primer sequences |
|---|---|
| 16S rDNA (mtDNA) | CGACTGTTTATCAAAAACAT |
| | CCGGTCTGAACTCAGATCACGT |
| Bpf1 | TCCTATCCTTGTAACTTCTCCAC* |
| | CGAAACCATGCAAATCAG* |
| Bpf2 | GCAGCTTCATTCAACATTCC* |
| | CATGATGGCTGAACAAG |
| Bpf4 | CCTCCGCCTCTATGAAG |
| | AAGCTGCGCTATCTGTGGAC |
| Bpf5 | TGTATGCTGACACTTAAAGAAACC* |
| | GGCTCCCAAATGAGGGTC |
| Bpf6 | CAAAAGTCACCGGCATAACC |
| | CTAGTTTTTGAGATGTAAACGG |
| Bpf8 | GGTTCCCATAGCATACAGTGC* |
| | CTCACTCTTTCACCAACG |
| Bpf10 | TGTCCAGCATGTCCAGTTC* |
| | CAGAGATGATATTGCAGTCAGG* |
| ITS1 | GTTTCCGTAGTGAACCTGCGG |
| | GTGCGTTCGAAGTGTCGATGATCAA |

The primer sequences are given from 5′ to 3′. The forward and reverse primers are on the first and second lines respectively. Primers for microsatellite loci (Bpf*i*) were either drawn from Charbonnel *et al*. (2000) (denoted by *) or redesigned from sequences (GenBank references in Charbonnel *et al*., 2000) for loci Bpf1, Bpf2, Bpf5, Bpf8 and Bpf10, and references AF189701 and AF189703 for loci Bpf4 and Bpf6). For the 16S rDNA, see Palumbi (1996).

analysis was conducted using Arlequin 2.0 (Schneider *et al*., 1999). The sequence diversity within any region was described using $\theta_K = K/a_n$ where $a_n = 1 + 1/2 + \cdots + 1/(n-1)$, $n$ being the sample size for that region.

A minimum-spanning haplotype network was inferred using Arlequin 2.0 for each locus. *B. alexandrina* and *B. sudanica* are closely related to *B. pfeifferi* (Campbell *et al*., 2000; DeJong *et al*., 2001), and were used as outgroups to infer the sequence of the most recent common ancestor (MRCA) at the origin of the sequences sampled from populations of *B. pfeifferi*. This was not possible for ITS1 because several mutations were indels in an AT-rich region which was difficult to align across species. Similarly sequences from *B. glabrata* were used to orient mutations along branches either to *B. pfeifferi* or to the *B. alexandrina–B. sudanica* group. This was possible for the mtDNA, Bpf1, Bpf2, Bpf6 and Bpf10. Neutrality was tested using the methods of Hudson *et al*. (1987; referred to below as the HKA test), Tajima (1989) and Fu (1997). For the latter two, the coalescent simulations were conditioned on $\pi$, and the associated statistics were calculated ($D$ and $Fs$ respectively). DnaSp 3.51 (Rozas & Rozas, 1999) and Arlequin 2.0 (Schneider *et al*., 1999) were used to conduct these tests. Neutrality and demographic processes were further tested using the method of Galtier *et al*. (2000) based on the coalescence theory and maximum likelihood (ML) estimation. This approach is more powerful than the statistics-based analyses mentioned above, especially with multilocus data. A series of hierarchical models of population history were fitted to the data. The simplest one (model 1) assumes a neutral, constant population size coalescent. $\theta$ was estimated for each locus ($p$ parameters for $p$ loci). Model 2 assumes a single bottleneck occurring at time $T$ (in $2N_e$ generations) in the past during a short period and of strength $S$ (technically $S$, measured in units $2N_e$ generations, is the time that would be required for an equal expected amount of coalescence if the population size had not changed). The population then recovers its initial size. This model has $p+2$ parameters since $\theta$ is estimated for each locus in addition to $S$ and $T$. Model 3 is similar to model 2 but it assumes that the mtDNA and the nuclear sequences underwent different bottlenecks. This model therefore has $p+4$ parameters, since $S$ and $T$ are independently estimated for the mtDNA and nuclear sequences. In model 4, $\theta$, $S$ and $T$ were estimated for each locus ($3p$ parameters) allowing for distinct selective sweeps. Likelihood-ratio tests were performed to compare these models. Twice the logarithm of the likelihood ratio of two competing models asymptotically follows a chi-square distribution with $j$ degrees of freedom where $j$ is the difference between the number of parameters of the two models. The models were adjusted to consider the mtDNA data only (models 3 and 4 are identical to model 2), nuclear DNA only (model 3 is equivalent to model 2), and both datasets. In the latter situation, the mating system was taken into account. As *B. pfeifferi* is a highly selfing, simultaneous hermaphroditic species, the effective sizes of nuclear and mitochondrial genes are expected to be identical. However, the sister species of *B. pfeifferi* are outcrossing hermaphrodites, meaning that *B. pfeifferi* switched towards self-fertilization at some point in its history. For model 2, we therefore conducted calculations in which the ratio of effective sizes of mitochondrial to nuclear genes varied between 1 and 0·5. This difference in effective size is not taken into account in the other models, since the effective size is directly adjusted to $\theta$ values.

## 3. Results

### (i) *AFLP data*

Amplifications produced 50 phylogenetically informative bands in *B. pfeifferi* populations, 29 fixed bands in *B. pfeifferi* populations that did not occur in *B. sudanica*, and 19 bands fixed in both species. Each population exhibited a specific multilocus profile (fingerprint). Six to nine bands were specific to each geographic region. The two phylogenetic methods

Table 3. *Characteristics of the nine sequenced DNA fragments in* B. pfeifferi

| Locus | Length | $n_a$ | $n_e$ | $\Pi$ | $\pi$ | $K$ | $D_{p/as}$ | $D_{a/s}$ | $D$ | $Fs$ | $\Phi_{st}$ | $F_{st}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16S RNA (mtDNA) | 393 | 5 | 3·46 | 1·750 (1·131) | 0·0045 (0·0034) | 7 | 15 (11 ; 4) | 6 (2 ; 4) | −1·674 (0·018) | −1·449 (0·083) | −0·124 | −0·026 |
| Bpf1 | 140 | 4 | 4·19 | 2·821 (1·891) | 0·0201 (0·0158) | 6 | 8 (2 ; 6) | 0 (0 ; 0) | 1·022 (0·858) | 0·869 (0·662) | 0·948* | 0·794* |
| Bpf2 | 278 | 3 | 2·30 | 0·679 (0·654) | 0·0024 (0·0023) | 2 | 5 (2 ; 3) | 0 (0 ; 0) | −0·448 (0·330) | −0·478 (0·167) | 0·259 | 0·259 |
| Bpf4 | 137 | 2 | 1·56 | 0·250 (0·194) | 0·0018 (0·0051) | 1 | 2 (nd) | 0 (0 ; 0) | −1·054 (0·201) | −0·182 (0·218) | −0·045 | −0·11 |
| Bpf5 | 312 | 4 | 3·49 | 1·785 (1·149) | 0·0057 (0·0042) | 5 | 5 (nd) | 0 (0 ; 0) | −0·335 (0·426) | −0·073 (0·434) | 0·783* | 0·286 |
| Bpf6 | 358 | 4 | 3·21 | 1·464 (0·987) | 0·0041 (0·0031) | 4 | 5 (3 ; 2) | 0 (0 ; 0) | −0·221 (0·431) | −0·470 (0·297) | 0·467* | 0·709* |
| Bpf8 | 321 | 6 | 4·13 | 2·714 (2·286) | 0·0085 (0·0070) | 7 | 6 (nd) | 2 (0 ; 2) | 0·025 (0·540) | −1·848 (0·068) | 0·416* | 0·043 |
| Bpf10 | 160 | 4 | 4·13 | 2·714 (1·608) | 0·0170 (0·0012) | 9 | 5 (2 ; 3) | 0 (0 ; 0) | −0·581 (0·302) | 0·787 (0·658) | −0·045 | 0·090 |
| ITS1 | 640 | 7 | 4·61 | 3·678 (3·263) | 0·0057 (0·0033) | 10 | 9 (nd) | 2 (1 ; 1) | −0·228 (0·428) | −2·679(0·047) | 0·063 | 0·223 |

Length is the size of sequenced fragments in base pairs, excluding the tandemly-repeated area for the microsatellite loci. $n_a$ is the number of alleles. $n_e$ is the expected number of alleles based on $\Pi$. $\Pi$ is the average difference between pairs of sequence and $\pi$ is the nucleotide diversity (both with standard deviation in parentheses). $K$ is the number of segregating sites. $D_{p/as}$ is the number of segregating sites between the ancestors of *B. pfeifferi* and *B. alexandrina–B. sudanica* sequences (when it was possible to root the tree with *B. glabrata* as an outgroup, the numbers of sites separating the root from the ancestor of *B. pfeifferi* and from the ancestor of *B. alexandrina–B. sudanica* respectively are given in parentheses; nd, not determined, because it was not possible to root the tree). $D_{a/s}$ is the number of segregating sites in the *B. alexandrina–B. sudanica* group (with the number of sites on the branches going to *B. alexandrina* and *B. sudanica* respectively in parentheses). $D$ and $Fs$ are Tajima's and Fu's statistics respectively with their probability in parentheses. $\Phi_{st}$ and $F_{st}$ are AMOVA statistics of population structure when the molecular difference among haplotypes was and was not taken into account respectively (* refers to estimates significantly different from zero).

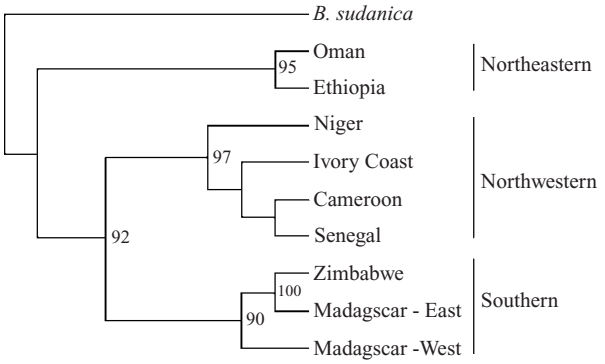


Fig. 2. Phylogenetic relationships among populations of *B. pfeifferi* reconstructed using AFLP data with the Dollo parsimony method. *B. sudanica* was used as an outgroup. The East and West populations of Madagascar are Irondro (see Table 1) and Andilamana respectively. Northeastern, Northwestern and Southern refer to the three groups of populations. Bootstrap values (in percent) supporting nodes higher than 90% are also indicated.

produced identical genealogies, with slight differences in bootstrap values. Each geographic region formed a strongly supported clade (bootstrap values $>0{\cdot}9$), suggesting isolation by distance (see Fig. 2). An interesting result is that the population from eastern Madagascar clustered with a population from Zimbabwe, rather than with that from western Madagascar. Further AFLP analyses including a larger number of eastern and western Malagasy populations confirmed this, and little variation was found within populations (data not shown). This is consistent with microsatellite data, and strongly supports the hypothesis of two independent introductions of *B. pfeifferi* in Madagascar (Charbonnel *et al.*, 2002*a*).

### (ii) *Variability of DNA sequences*

All sequences except Bpf4, Bpf10 and ITS1 showed a significant excess of AT nucleotides (Fisher exact test; $P<3\times10^{-4}$). Searches for similarity performed in the GenBank database using BLAST tools (Altschul *et al.*, 1997) did not find similarities with the microsatellite flanking regions. Patterns of variability are reported in Table 3 and Fig. 1. Five haplotypes were detected at the mtDNA locus in the eight *B. pfeifferi* populations. Additional screening showed the same haplotypes in other populations from Cameroon, Ivory Coast, Zimbabwe (one population per country) and Madagascar (four populations). The haplotypes were separated by no more than four mutations, and the nucleotide diversity per site was quite low ($\pi=$ 0·0045). The sequence of microsatellite flanking regions showed two to six alleles per locus over all populations. Values of molecular diversity are given in Table 3. The number of segregating sites ranged from one (Bpf4) to nine (Bpf10) and up to eight mutations

https://doi.org/10.1017/S0016672303006219 Published online by Cambridge University Press

distinguished two sequences at a given locus (Bpf8). The nucleotide diversity was lower at three loci (Bpf2, Bpf4 and Bpf6) than that for mtDNA, but for Bpf1 was as high as 0·0202. The average nucleotide diversity over loci was 0·0062. The ITS1 sequence was the most polymorphic locus, with seven alleles, ten polymorphic sites and up to seven mutations separating haplotypes. However, its nucleotide diversity per site is moderate (0·0057).

(iii) *Phylogenetic reconstructions*

All minimum-spanning haplotype networks were well resolved (see Fig. 1), with the exception of the genealogy of Bpf8, which implied a mutation occurring twice. The MRCA sequence occurred in the populations analysed at all loci except Bpf8, and had the most widespread geographic distribution at all loci except Bpf1 and Bpf6. An interesting result is that the number of mutations separating haplotypes from the MCRA (at all loci) was very low for the southern and northeastern populations, varying between five for a Madagascar population and seven for both Oman and Ethiopia (Fig. 1). In the western zone, the Niger population was separated from the MRCA by six mutational events, while there were nine, 15 and 17 mutations for the Senegal, Ivory Coast and Cameroon populations respectively.

Seventy fixed sites were found from *B. pfeifferi* MRCA to *B. alexandrina* and *B. sudanica* (summing over values of $D_{p/as}$ and $D_{a/s}$ in Table 3), 10 of which are fixed in either *B. alexandrina* or *B. sudanica*. Twenty-one and 49 of these sites were detected at the mtDNA (over 393 bp) and nuclear (over 2346 bp) loci respectively. Comparing segregating sites within *B. pfeifferi* and fixed sites between *B. pfeifferi* MRCA and *B. alexandrina*/*B. sudanica* ancestor ($K$ and $D_{p/as}$ in Table 3 respectively) at the mtDNA and nuclear loci produced a non-significant difference ($G$-test: $G=2{\cdot}26$, $P=0{\cdot}13$; HKA test: $\chi^2=0{\cdot}21$, $P=0{\cdot}65$). Rooting the tree with *B. glabrata* as outgroup was possible for five loci (mtDNA, Bpf1, Bpf2, Bpf6 and Bpf10), and allowed us to position mutations along branches going from *B. pfeifferi* MRCA to the ancestor of *B. pfeifferi* and the *B. alexandrina–B. sudanica* group. Eleven and eight mutations were fixed in these branches for the mtDNA and nuclear loci respectively, while seven and 21 mutations were segregating in *B. pfeifferi* respectively. The statistical results were contrasted ($G$-test: $G=8{\cdot}53$, $P=0{\cdot}004$; HKA test: $\chi^2=0{\cdot}45$, $P=0{\cdot}50$). There is therefore mixed evidence that the MRCA sequence for the mtDNA is recent compared with those at the nuclear loci.

(iv) *Population structure*

The distribution of haplotypes among *B. pfeifferi* populations is given in Fig. 1. The most frequent mtDNA haplotype was found in Niger, Ethiopia, Oman and Zimbabwe, and the four variants were each found in a single population. The genealogy of alleles at microsatellite loci was generally more consistent with the geographical location of populations, i.e. that geographically close populations were also phylogenetically so. The most striking example is the Bpf1 locus, for which each of the three geographic regions could be distinguished on the basis of two to four mutations. In addition, apomorphic mutations discriminated populations from the western group (Bpf8) from those in the northeastern (Bpf5) and the southern (Bpf6) groups. A more ambiguous situation was found at ITS1: the Cameroon and Niger populations were most closely related to the northeastern populations, while the Ivory Coast and Senegal populations were most closely related to the southern populations.

The molecular diversity among populations within regions was very low for both the southern and northeastern groups, with only three variable loci ($\theta_K=5$; Fig. 1). This contrasts with the pattern observed in the western region where all loci, except Bpf1, were variable ($\theta_K=9{\cdot}27$). A large fraction of this diversity could be attributed to Bpf10, which had nine out of the total of 16 polymorphic sites, and was variable in the western region only. A similar pattern was observed at ITS1 ($\theta_K=4{\cdot}36$, 0·00 and 2·00 in the western, southern and northern populations respectively). The AMOVAs indicated significant $\Phi_{st}$ and $F_{st}$ values at four and two nuclear loci only (Table 3), and no structure was detected at the mtDNA locus. There is therefore partial indication only of geographic structure, although the tests might be weak given the sample size.

(v) *Neutrality and demographic equilibrium*

The results of the neutrality tests are reported in Table 3. Tajima's (1989) test rejects the neutral hypothesis for the mtDNA locus only, due to an excess of recent mutations. Other tests of neutrality (Fu & Li, 1993) and the associated statistics ($D$ and $Fs$) gave essentially identical results (not shown). Fu's (1997) test was significant for the ITS1 locus only, indicating an excess of haplotypes. Negative values of $D$ and $Fs$ were also observed at the mtDNA and Bpf8 loci, with $P$ values close to 0·05. Note that conditioning coalescent simulations on $S$ in these tests (see Rozas & Rozas, 1999) returned slightly higher $P$ values, pushing to a cautious interpretation of results (see Discussion).

The results of the Galtier *et al.*'s (2000) method are reported in Table 4. Model 2 (bottleneck) was significantly better than model 1 (mutation–drift equilibrium) when only the mtDNA sequence was considered ($P=7\times10^{-3}$), suggesting a marked reduction in

https://doi.org/10.1017/S0016672303006219 Published online by Cambridge University Press

Table 4. *Results of the test* (*Galtier* et al.*, 2000*) *on bottlenecks and selective sweeps*

| Model | Parameters | mtDNA only | nDNA only | All loci |
|---|---|---|---|---|
| 1 | L | −7·72 | −73·34 | −81·06 |
| | Θ | 4·8 | 0·3 to 5·4 | 0·3 to 5·4 |
| 2 | L | −2·70* | −69·33* | −75·08 to −76·54 |
| | Θ | 90·1 | 0·4 to 5·6 | 0·5 to 7·6 |
| | $T$ | 0·018 | 0·80 | 0·55 to 0·66 |
| | $S$ | 4·71 | 4·36 | 2·90 to 4·00 |
| 3 | L | As model 2 | As model 2 | −72·03* |
| | Θ | | | 0·4 to 90·1 |
| | $T_{mt}$; $T_n$ | | | 0·018; 0·8 |
| | $S_{mt}$; $S_n$ | | | 4·71; 4·36 |
| 4 | L | −2·70 | −68·36 | −71·06 |
| | Θ | 90·1 | 0·30 to 9·5 | 0·30 to 90·1 |
| | $T$ | 0·018 | 0·38 to 1·80 | 0·018 to1·80 |
| | $S$ | 4·71 | 1·37 to 8·0 | 1·37 to 8·10 |

The table gives the likelihood (L) of tests, as well as the estimates of Θ, $T$ and $S$. The tests were performed either for the mtDNA sequence, for nuclear loci, or for all loci. The models are described in the Materials and methods section. Tests were based on log-likelihood ratios, and * denotes the (significantly) best model. The associated $P$ values are reported in the text. The 'mt' and 'n' subscripts refer to estimates for mitochondrial and nuclear data. Note that, in order to keep this table at a reasonable size, a range of values is given when multiple estimates of some parameters were produced, either across loci (third and fourth columns) or across values of the ratio of effective size of mtDNA and nDNA (fourth column, model 2).

diversity. A similar result was found when only the nuclear loci only were analysed ($P=0·02$). Moreover, model 3 (distinct sweep for each locus) was not significantly better than model 2 for nuclear loci ($P=1$), suggesting that the marked reduction in diversity at these loci was due to a demographic event. When loci from both genomes were considered in the same analysis, the best model was model 3 ($P<0·01$ for models 2 versus 1; $P<0·03$ for models 3 versus 2; $P=1$ for models 4 versus 3). This indicates that distinct events of diversity reduction have affected the nuclear and the mitochondrial genomes. This does not reflect the smaller effective size of the mtDNA, as the result is the same whatever the ratio of effective population sizes assumed in model 2. This means that even if *B. pfeifferi* switched from outcrossing to selfing quite recently, this has limited influence on our results. The estimate of $T$ for the mtDNA sequence was much lower than for the nuclear loci, in agreement with this result, suggesting a more recent event affecting mtDNA than nuclear gene diversity. Note that as the Galtier *et al*. (2000) method was applied to three datasets (mtDNA only, nuclear DNA only, combined data), the significance value should not be compared directly across columns of Table 4. For example, the significant fit of model 3 to the combined data anticipates that of model 2 to genome-specific data.

## 4. Discussion

### (i) *Limited variability within populations*

Data obtained at 10 microsatellite loci showed limited variation within the populations studied here (B. Angers, unpublished data), and this was the very reason for sequencing one to four haplotypes (that is individuals, because of high homozygosity) per population. No variation was indeed found when several haplotypes were sequenced per population. This is consistent with results obtained in more than 50 populations sampled all over Madagascar (Charbonnel *et al*., 2002*a*, *b*). Similar results have also been obtained in *Bulinus truncatus*, another African freshwater snail with a natural history very similar to that of *B. pfeifferi* (Viard *et al*., 1997*a*, *b*). There are two likely explanations for the low genetic variation: (i) Self-fertilization reduces the effective size of populations ($N_e$) by up to a factor 2, and enhances the influence of indirect selection through either hitchhiking, or background selection (see Charlesworth *et al*., 1993; Jarne, 1995). (ii) These populations probably have a limited effective size as a consequence of regular or stochastic fluctuations in individual density in response to environmental processes (climatic or ecological; Charbonnel *et al*., 2002*a*, *b*; review in Städler & Jarne, 1997). The main genetic consequence is the reduction in the effective size to about its harmonic mean over generations (Hartl & Clark, 1997).

### (ii) *Genetic diversity and population history*

Collectively our results suggest that the geographical distribution of genetic diversity in *B. pfeifferi* in Africa results from dispersion from Western Africa, where diversity is highest, involving a single (or a few) founding event. This conclusion is drawn from (i) the geographic distribution of diversity among groups of populations and (ii) its deviation from the neutral expectation

https://doi.org/10.1017/S0016672303006219 Published online by Cambridge University Press

at mutation–drift equilibrium. This interpretation might be affected by current population structure, as discussed below. (i) The AFLP data indicate that populations group according to geographic proximity (regions), suggesting that populations within regions have a common evolutionary history. Sequence data at four nuclear loci support the geographic pattern revealed through AFLP, since it shows high molecular variance among regions, and suggests some form of isolation by distance. However, some loci did not support this pattern (Bpf10 and ITS1) or reveal no structure (Bpf2 and Bpf4). This is not surprising as population structure is expected to generate substantial variance among loci (Vitalis *et al*., 2001 and references therein). A similar geographic pattern for the ITS locus can be inferred from the studies of Campbell *et al*. (2000) and DeJong *et al*. (2001). Furthermore, these authors detected only one extra-sequence (from Sudan) by comparison with the sequences reported here. This incongruence is not incompatible with a dispersion scenario from a genetically variable source. Large variance in allelic frequencies is expected to result from genetic drift after founder events. (ii) The tests of neutrality (Tajima, 1989; Fu & Li, 1993; Fu, 1997) did not indicate significant departures from mutation–drift equilibrium at any nuclear loci when they were analysed separately. An exception was the ITS1 locus when using Fu's (1997) test. However, the multilocus, coalescence-based maximum likelihood method of Galtier *et al*. (2000) detects a reduction in variability that is consistent with a single bottleneck event. Modelling independent selection events separately for the nuclear loci did not significantly improve the likelihood of the model. It remains possible that population structure, although not detected at all loci, affects the results of this test: the procedure was designed for a single population, and the influence of population structure has not been evaluated. Intuitively, population structure is expected to produce coalescent trees with long terminal branches, especially when migration is limited (large number of mutations; Nordborg, 2001) – the opposite of the expectations for a bottlenecked population (short terminal branches). However, specific models of population structure may presumably create coalescent topologies similar to the ones observed here (see e.g. Wilkins & Wakeley, 2002). More theoretical work needs to be done in this area. Note also that population structure was not detected by neutrality tests, but they are notoriously conservative. A further complication is the mating system: if *B. pfeifferi* switched recently from outcrossing to selfing, the ratio of effective size of mtDNA and nuclear DNA is no longer 1. However, we showed that this does not affect the results of the method of Galtier *et al*. (2000).

Where does *B. pfeifferi* come from? The amount of diversity among groups of populations suggests that *B. pfeifferi* extended its distribution from western Africa, and colonized the southern and northeastern regions. When considering all populations at once, substantial variation was detected across loci in the amount of variability (a factor of 20 in nucleotide diversity between the least and the most polymorphic loci). However, far more variation was detected in the western group than in the two other groups, pleading for a West African origin of *B. pfeifferi*. An interesting element is the wide distribution of the MRCA, assuming that it has been properly inferred. This is not surprising since older alleles are expected to have a more widespread distribution than younger ones. MRCAs are also expected to occur more frequently than other alleles among individuals founding new populations, which may explain their high frequency in the southern and northeastern regions compared with the western 'source' (but see Cann *et al*., 1987).

(iii) *Selection on the mitochondrial molecule?*

Two surprising results regarding the mitochondrial gene studied are the rather low variability (e.g. $\pi =$ 0·0045 versus an average of 0·0062 for the other loci) and the near absence of population structure, with the MRCA found throughout Africa. These results were unexpected given the large geographical area surveyed and results reported from other species of pulmonate snails (Ross, 1999; Shimizu & Ueshima, 2000; but see Pfenninger & Magnin, 2001). They also contrast with those obtained at the nuclear sequences analysed here (see discussion above). De Jong *et al*. (2001) independently analysed the same mtDNA locus in populations from Madagascar, Senegal, Cameroon and Sudan (GenBank Accession No. AY030193-96). No additional variation was revealed, confirming that the pattern here is not simply due to small sample size. It would certainly be worthwhile sequencing extra mtDNA genes to check whether the current pattern extends beyond the 16S rDNA region.

A first obvious explanation might be that the mtDNA gene has a low mutation rate. However, this runs counter to what is generally found (e.g. Graustein *et al*., 2002 in *Caenorhabditis* species), and is ruled out by the higher divergence between *B. pfeifferi* and the *B. alexandrina–B. sudanica* group for the mtDNA gene than for most of the nuclear loci (except for Bpf1). Similarly, the mtDNA divergence between *B. alexandrina* and *B. sudanica* is higher than that at any nuclear locus. A second explanation is genetic homogenization resulting from high migration rates among populations. This is not impossible in a species in which long-distance dispersal may be mediated by birds or human activities (Brown, 1994; Viard *et al*., 1997*a*). However, the regional structure observed at both the AFLP and nuclear sequence data strongly

argues against this interpretation. Taking into account both the rather high mtDNA mutation rate and the marked population structure at nuclear loci, the third, most likely, explanation is that the mtDNA molecule experienced a marked reduction in ancestral polymorphism. This is supported by all neutrality tests, with the exception of the HKA test.

The analysis based on the framework developed by Galtier *et al*. (2000) reveals that a single bottleneck event alone cannot explain the reduction in variability at both the mtDNA gene and nuclear loci. A striking result is the almost 50-fold difference between nuclear and mitochondrial genes in the estimate of the time since the bottleneck (parameter *T*). Certainly, time estimates with maximum-likelihood methods may have large variances, and deriving standard errors from the Griffiths-Tavaré algorithm is uneasy, because the shape of the distribution of the likelihood across genealogies is unknown. However, the simulations conducted by Galtier *et al*. (2000; their table 2) suggest that the size of our dataset allows discrimination between old and recent (selective sweep) bottlenecks. Importantly, the variance is taken into account in the likelihood ratio test. If the mitochondrial data were compatible with a bottleneck as long ago as the bottleneck detected from the nuclear data, then model 2 should be favoured (i.e. the difference in log-likelihood of the mitochondrial data for 'recent *T*' versus 'old *T*' should be low). There are three possible explanations for this result: population structure, purifying selection and selective sweep.

We mentioned above that population structure may interfere in the neutrality tests when looking for bottlenecks, since it tends to increase the variance among loci in diversity (Vitalis *et al*., 2001) and might produce coalescent trees mimicking those trees obtained under selection (Wilkins & Wakeley, 2002). It is therefore likely that population structure affects neutrality tests, in an unknown fashion though, as has been suggested for the HKA test (Nielsen, 2001). Population structure might therefore account for the difference observed at the mtDNA locus. However, it would be surprising if the only locus which is a potential target of selection in this study were the one which is thrown out of the distribution in neutrality tests because of population structure.

Williamson & Orive (2002) pointed out that purifying selection modifies coalescent trees in a way similar to what can be expected under demographic bottlenecks or selective sweeps. Using simulations, these authors showed that it might have a strong enough effect to affect the outcome of classical neutrality tests (presumably also that of the Galtier *et al*. (2000) method), mimicking the effect of positive selection. This is true when the selective coefficients and mutation rates are large enough, and when mutation effects are uniformly distributed across mutation events (Williamson & Orive, 2002). The question, of course, is whether these conditions hold in the actual world. Evaluating the two first conditions is uneasy, but there is evidence from studies on mutation effects that the last one is unlikely to be met (see e.g. Shaw *et al*., 2002). Clearly purifying selection cannot be rejected, but this calls for more work evaluating its actual consequences under the conditions studied here.

The last explanation is a selective sweep on the mtDNA. Several empirical results have indeed suggested that positive selection markedly influences molecular diversity (e.g. Begun & Aquadro, 1991; Schlötterer *et al*., 1997; Nurminsky *et al*., 2001). However, it has turned out to be difficult to disentangle the consequences of the different mechanisms involved in diversity reduction, including selective sweeps, background selection and demographic events. A solution that has been adopted in a few studies, including the current one, is to compare patterns of diversity across several genes from the same genome. Nurminsky *et al*. (2001) provided a good example, comparing the patterns of variation at a series of about 15 genes between *Drosophila melanogaster* and *Drosophila simulans*. A newly evolved gene in *D. melanogaster*, which does not exist in *D. simulans*, has apparently been associated with a local loss of variation through positive selection. However, very few studies, if any, indicate that the mtDNA molecule experienced positive selection (but see Ballard, 2000), previous studies reporting either purifying selection or no selection at all (reviewed in Weinreich & Rand, 2000; Gerber *et al*., 2001). These conclusions are largely based on McDonald–Kreitman tests (McDonald & Kreitman, 1991), which may be conservative with regard to selective sweeps (Gerber *et al*., 2001), and might therefore have to be reconsidered.

Whether the pattern is observed at the mtDNA due to population structure or selection, an important conclusion of our study is that evolutionary studies should not be based on the single mtDNA molecule, as already highlighted by several authors.

We thank P. Brémond, S. Chandiwana, B. Delay, J.-P. Pointier, D. Rollinson and J. Webster for providing snails or help with sampling, C. Debain and the 'Service commun des marqueurs génétiques' of Centre d'Etude Fonctionnelle et Evolutive for help in the molecular analysis, and P. Capy, D. Charlesworth, C. Jones, T. Lenormand, M. Orive, S. Otto and an anonymous reviewer for constructive comments (including linguistic ones) on the manuscript. This work was supported by the Centre National de la Recherche Scientifique 'Jeunes équipes' program and by a grant from the Ministère de la Recherche (program 'Microbiologie et maladies infectieuses et parasitaires; 95') to P.J. B.A. was supported by fellowships from the Centre National de la Recherche Scientifique (poste de chercheur associé) and the Fond pour la Formation de Chercheurs et l'Aide à la Recherche, and N.C. by a fellowship from the Ministère de la Recherche.

## References


Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Avise, J. C. (2000). *Phylogeography. The History and Formation of Species*. Harvard: Harvard University Press.

Ballard, J. W. O. (2000). Comparative genomics of mitochondrial DNA in *Drosophila simulans*. *Journal of Molecular Evolution* **51**, 64–75.

Bandoni, S. M., Mulvey, M., Koech, D. K. & Loker, E. S. (1990). Genetic structure of Kenyan populations of *Biomphalaria pfeifferi* (Gastropoda: Planorbidae). *Journal of Molluscan Studies* **56**, 383–391.

Begun, D. J. & Aquadro, C. F. (1991). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow-achaete* region. *Genetics* **129**, 1147–1158.

Boissinot, S. & Boursot, P. (1997). Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? *Genetics* **146**, 1019–1034.

Brown, D. S. (1994). *Freshwater Snails of Africa and their Medical Importance*. London: Taylor & Francis.

Campbell, G., Jones, C. S., Lockyer, A. E., Hughes, S., Brown, D. S., Noble, L. R. & Rollinson, D. (2000). Molecular evidence supports an African affinity of the Neotropical freshwater gastropod, *Biomphalaria glabrata*, Say 1818, an intermediate host for *Schistosoma mansoni*. *Proceedings of the Royal Society of London, Series B* **267**, 2351–2358.

Cann, R. L. (2001). Genetic clues to dispersal in human populations: retracing the past from the present. *Science* **291**, 1742–1748.

Cann, R. L., Stonekind, M. & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Charbonnel, N., Angers, B., Razatavonjizay, R., Brémond, P. & Jarne, P. (2000). Microsatellite variation in the freshwater snail *Biomphalaria pfeifferi*. *Molecular Ecology* **9**, 1006–1007.

Charbonnel, N., Angers, B., Rasatavonjizay, R., Brémond, P., Debain, C. & Jarne, P. (2002*a*). The influence of mating system, demography, parasites and colonization on the population structure of *Biomphalaria pfeifferi* in Madagascar. *Molecular Ecology* **11**, 2213–2228.

Charbonnel, N., Angers, B., Razatavonjizay, R., Brémond, P. & Jarne, P. (2002*b*). Exploring the evolutionary aspects of the complex metapopulation dynamics in *Biomphalaria pfeifferi*, the intermediate host of *Schistosoma mansoni*. *Journal of Evolutionary Biology* **15**, 248–261.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

DeJong, R. J., Morgan, J. A. T., Paraense, W. L., Pointier, J.-P., Amarista, M., *et al*. (2001). Evolutionary relationships and biogeography of *Biomphalaria* (Gastropoda: Planorbidae) with implications regarding its role as host of the human bloodfluke, *Schistosoma mansoni*. *Molecular Biology and Evolution* **18**, 2225–2239.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Farris, J. S. (1977). Phylogenetic analysis under Dollo's law. *Systematic Zoology* **26**, 77–88.

Felsenstein, J. (1994). PHYLIP (Phylogeny Inference Package).

Fu, Y.-X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Fu, Y.-X. & Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Galtier, N., Depaulis, F. & Barton, N. H. (2000). Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**, 981–987.

Gerber, A. S., Loggins, R., Kumar, S. & Dowling, T. E. (2001). Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial evolution? *Annual Review of Genetics* **35**, 539–566.

Goldstein, D. B. & Schlötterer, C. (ed.) (1999). *Microsatellites. Evolution and Applications*. Oxford: Oxford University Press.

Graustein, A., Gaspar, J. M., Walters, J. R. & Palopoli, M. F. (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**, 99–107.

Hare, M. P. (2001). Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* **16**, 700–706.

Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R. & Sherry, S. T. (1998). Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the USA* **95**, 1961–1967.

Hartl, D. L. & Clark, A. G. (1997). *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates.

Hey, J. & Harris, E. (1999). Population bottlenecks and patterns of human polymorphism. *Molecular Biology and Evolution* **16**, 1423–1426.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Jarne, P. (1995). Mating system, bottlenecks and genetic polymorphism in hermaphroditic animals. *Genetical Research* **65**, 193–207.

Kluge, A. G. & Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **18**, 1–18.

Li, W. H. (1997). *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Lynch, M. & Milligan, B. G. (1994). Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**, 91–99.

McDonald, J. H. & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654.

Mimpfoundi, R. & Greer, G. J. (1990). Allozyme variation among populations of *Biomphalaria pfeifferi* (Krauss, 1848) (Gastropoda: Planorbidae) in Cameroon. *Journal of Molluscan Studies* **56**, 461–467.

Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**, 641–647.

Nordborg, M. (2001). Coalescent theory. In *Handbook of Statistical Genetics* (ed. D. J. Balding, M. J. Bishop & C. Cannings), pp. 179–212. Chichester: John Wiley.

Nurminsky, D., De Aguilar, D., Bustamante, C. D. & Hartl, D. L. (2001). Chromosomal effects of rapid gene evolution in *Drosophila melanogaster*. *Science* **291**, 128–130.

Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer and Applied Biosciences* **12**, 357–358.

Palumbi, S. R. (1996). Nucleic acids. II. The polymerase chain reaction. In *Molecular Systematics* (ed. D. M. Hillis, C. Moritz & B. K. Marble), pp. 205–248. Sunderland, MA: Sinauer Associates.

Pannell, J. R. & Charlesworth, B. (2000). Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London, Series B* **355**, 1851–1864.

Pfenninger, M. & Magnin, F. (2001). Phenotypic evolution and hidden speciation in *Candidula unifasciata* ssp. (Helicillenae, Gastropoda) inferred by 16S variation and quantitative shell traits. *Molecular Ecology* **10**, 2541–2554.

Ross, T. K. (1999). Phylogeography and conservation genetics of the Iowa Pleistocene snail. *Molecular Ecology* **8**, 1363–1373.

Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Schlötterer, C., Vogl, C. & Tautz, D. (1997). Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* **146**, 309–320.

Schneider, S., Kuffer, J. M., Roessli, D. & Excoffier, L. (1999). Arlequin v.1.1, user manual. University of Geneva, Switzerland.

Shaw, F. H., Geyer, C. J. & Shaw, R. G. (2002). A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**, 453–463.

Shimizu, Y. & Ueshima, R. (2000). Historical biogeography and interspecific mtDNA introgression in *Euhadra peliomphala* (the Japanese land snail). *Heredity* **85**, 84–96.

Singh, R. S. & Krimbas, C. B. (ed.) (2000). *Evolutionary Genetics. From Molecules to Morphology*. Cambridge: Cambridge University Press.

Slatkin, M. & Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**, 603–613.

Städler, T. & Jarne, P. (1997). Population biology, genetic structure, and mating system parameters in freshwater snails. In *Ecology and Evolution of Freshwater Organisms* (ed. B. Streit, T. Städler & C. M. Lively), pp. 231–262. Basel: Birkhäuser.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Takahata, N., Lee, S.-H. & Satta, Y. (2001). Testing multiregionality of modern human origins. *Molecular Biology and Evolution* **18**, 172–183.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Valsecchi, E. (1998). Tissue boiling: a short-cut in DNA extraction for large-scale population screenings. *Molecular Ecology* **7**, 1243–1245.

Van Damme, D. (1984). *The Freshwater Mollusca of Northern Africa. Distribution, Biogeography and Palaeoecology*. Dordrecht: W. Junk.

Viard, F., Justy, F. & Jarne, P. (1997*a*). The influence of self-fertilization and bottlenecks on the genetic structure of subdivided populations: a case study using microsatellite markers in the freshwater snail *Bulinus truncatus*. *Evolution* **51**, 1518–1528.

Viard, F., Justy, F. & Jarne, P. (1997*b*). Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. *Genetics* **146**, 973–982.

Vitalis, R., Dawson, K. & Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811–1823.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., *et al*. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Weinreich, D. M. & Rand, D. M. (2000). Contrasting patterns of nonneutral evolution in proteins encoded in nuclear and mitochondrial genomes. *Genetics* **156**, 385–399.

Wilkins, J. F. & Wakeley, J. (2002). The coalescent in a continuous, finite, linear population. *Genetics* **161**, 873–888.

Williamson, S. & Orive, M. E. (2002). The genealogy of a sequence subject to purifying selection at multiple sites. *Molecular Biology and Evolution* **19**, 1376–1384.

Woolhouse, M. E. J. (1992). Population biology of the freshwater snail *Biomphalaria pfeifferi* in the Zimbabwe highveld. *Journal of Applied Ecology* **29**, 687–694.