Microsatellite Size Homoplasy, SSCP, and Population Structure: A Case Study in the Freshwater Snail *Bulinus truncatus*

B. Angers,* A. Estoup,†1 and P. Jarne*

*Centre d’Ecologie Fonctionnelle et Evolutive–Centre National de la Recherche Scientifique, Montpellier, France; and †Modélisation et Biologie Evolutive, URLB–INRA, Montpellier, France*

The extent of microsatellite size homoplasy, as well as its effect on several population genetics statistics, was investigated in natural populations using the single-strand conformation polymorphism (SSCP) method. The analysis was conducted using 240 individuals from 13 populations of the freshwater snail *Bulinus truncatus* at a GT,CTm compound microsatellite locus. We showed that SSCP can be used to uncover, at least partly, size homoplasy in the core sequence of this category of loci. Eight conformers (SSCP variants) were detected among the three size variants (electromorphs). Sequencing revealed that each conformer corresponded to a different combination of repeats in the GT, and CTm arrays. Part of this additional variability was detected within populations, resulting in a substantial increase in gene diversity in four populations. Additional variability also changed the values of parameters used to analyze population differentiation among populations: pairwise tests of differentiation were significant much more often with conformers than with electromorphs. On the other hand, pairwise estimates of Fst were either smaller or larger with conformers than with electromorphs, depending on whether or not electromorphs were shared among populations. However, estimates of Fst (or analogs) over all populations were very similar, ranging between 0.66 and 0.75. Our results were consistent with the theoretical prediction that homoplasy should not always lead to stronger population structure. Finally, conformer sequences and electromorph size distribution suggested that single-point and/or stepwise mutations occurring simultaneously in the different repeated arrays of compound microsatellites produce sequence variation without size variation and hence generate more size homoplasy than expected under a simple stepwise mutation model.

**Introduction**

Homoplasy is defined as apparent similarity masking evolutionary differences. For a genetic character, it occurs when different copies of a locus are identical in state but are not identical by descent. The occurrence of homoplasy is related to the way mutation produces new alleles. For example, it is expected under both the stepwise mutation model (SMM; Ohta and Kimura 1973) and the K-allele model (Kimura 1968), while it is not expected under the infinite-alleles model (IAM; Crow and Kimura 1964). Homoplasy depends not only on the mutation model, but also on the mutation rate (Estoup and Angers 1998). While homoplasy is a recurrent theme in molecular evolution and phylogeny, it has also recently attracted the attention of population geneticists, as a consequence of the rise of microsatellite loci as genetic markers. Microsatellites indeed mutate in a stepwise fashion, even if it is clear that the mutation process is often more complex than that represented by the various versions of the SMM (Macaubas et al. 1997; Colson and Goldstein 1999). However, a simple SMM has been chosen for convenience when building population genetic statistics and is supposed to be more adapted to microsatellite variation than the classical IAM (reviews by Jarne and Lagoda 1996; Estoup and Angers 1998). A nonexhaustive list of such statistics includes the genetic distance of Goldstein et al. (1995) and several analogs of Wright’s Fst (Slatkin 1995; Michalakis and Excoffier 1996; Rouset 1996).

Microsatellite alleles generally refer to DNA fragments of different sizes (electromorphs) revealed by classical electrophoretic methods. Electromorphs are identical in state (i.e., have identical sizes) but are not necessarily identical by descent. A fraction of this homoplasy may be uncovered because a given electromorph may hide different sequences. This corresponds to the molecularly observable fraction of size homoplasy and will hereinafter be referred to as size homoplasy. Compound or interrupted microsatellite loci, or loci showing variation in their flanking regions, constitute favorable situations for the detection of size homoplasy (Estoup et al. 1995; Garza and Freimer 1996; Grimaldi and Crouau-Roy 1997). It is worth noting that an unknown part of homoplasy at such microsatellite loci remains concealed, because even identical sequences may not be identical by descent. The few empirical studies which attempted to assess the effect of size homoplasy on population genetics parameters suggest that uncovering size homoplasy may substantially affect the estimation of several of these parameters (Estoup et al. 1995; Viard et al. 1998). However, these studies were limited in the number of individuals analyzed because detection of size homoplasy was based on cloning and sequencing of electromorphs. In this context, a proper population genetics study should make use of a method allowing (1) the ability to distinguish between electromorphs and sequences and (2) to analysis of a sufficiently large number of individuals per population (ca. 15–30) at a low cost in both time and money. The single-strand conformation polymorphism (SSCP; Orita et al. 1989) method has these qualities. It relies on the
ability of a single nucleotide change to alter the electrophoretic mobility of pieces of DNA and allows analysis of both copies of a given gene for a given individual. SSCP variants will be referred to as conformers. Orti, Pearse, and Avise (1997) showed that the SSCP technique can be used for detecting mutations along the flanking regions and for assessing size convergence at a microsatellite locus. However, these authors used a pure-repeat locus and hence could not detect size homoplasy occurring in the microsatellite core region, where most length variability arises. Moreover, they were not concerned with the effect of size homoplasy on population genetics parameters.

In this paper, we investigate the extent of microsatellite size homoplasy in natural populations of the selving snail {Bulinus truncatus} using the SSCP method and evaluate its effects on the estimation of several population genetics parameters. Size homoplasy was studied at the compound microsatellite locus BT1 (Viard et al. 1996). This locus was chosen for two reasons: (1) previous work based on sequencing revealed that electromorphs may hide several sequences (Viard et al. 1998), and (2) length variation at this locus is relatively limited, and a given electromorph may be shared by several populations, allowing the possibility of substantial size homoplasy (Viard et al. 1996; Viard, Justy, and Jarne 1997). The SSCP screening was performed for 13 populations sampled over a large part of the distribution area of {B. truncatus}, and data analysis was conducted using some standard and more recent tools.

### Materials and Methods

The analysis was conducted using individuals from 13 populations of the freshwater snail {B. truncatus} originating from seven different countries (table 1). Details about the location and genetic structure of these populations can be found in Viard, Justy, and Jarne (1997). Snails from these populations mostly self-fertilize and are generally homozygous at all loci studied (Viard, Doums, and Jarne 1997). The focus here was on the compound locus BT1, characterized by a GT₁CTₐ core sequence (Jarne et al. 1994). A previous study on size homoplasy indicated that some populations display electromorphs with two different sequences at this locus (table 1; Viard et al. 1998). The electromorphs studied here were electromorphs 180, 184, and 186, which correspond to 13, 15, and 16 repeats, respectively. There was no additional electromorph in the populations studied. Two to 35 electromorph copies per population were studied, for a total of 480 copies (table 1).

Preliminary SSCP tests were performed on individuals known to carry different sequences of a given electromorph, as previously observed by Viard et al. (1998). This allowed us to set up SSCP experimental conditions that discriminated unambiguously among copies of electromorphs with similar sizes and different sequences based on their electrophoretic mobility. The BT1 locus was amplified using radioactive labeling as described by Viard et al. (1996), except that the BT1–2 primer was labeled. Discrimination of conformers using the SSCP technique was accomplished as follows: 1 μl of amplified DNA was mixed with 5 μl of denaturing loading buffer (deionized formamide and 0.01% xylene cyanol). Samples were denatured for 2 min at 95°C and kept on ice until gel loading. They were run in a 6% polyacrylamide gel (37.1 acrylamide to methylbisacrylamide) with 1 × TBE buffer on a vertical electrophoresis system at constant power (20 W) and temperature (4°C) for 6 h. DNA sequencing was performed in order to confirm that the conformers observed in the present study differed in sequence. Sequencing of conformers was performed on purified PCR products (Wizard, Promega) using the ftml DNA sequencing system (Promega).

Data analysis was conducted using either conformers or electromorphs, that is, considering or not considering the fraction of size homoplasy revealed by SSCP.
The genetic variability within populations was estimated by counting the number of electromorphs or conformers and by computing the observed heterozygote frequency and gene diversity (Nei 1987). An estimate of Wright’s inbreeding coefficient $F_{st}$ was computed according to Weir and Cockerham (1984). This allowed the estimation of the mean population selfing rate ($\hat{S}$) using the formula $\hat{S} = 2\hat{f}(1 + \hat{f})$ (Pollak 1987). Exact tests were performed to test for genotypic linkage disequilibria between the BT1 locus and three other polymorphic loci (Viard, Justy, and Jarne 1997) and for differentiation among pairs of populations. The significance of tests was adjusted using a sequential Bonferroni procedure (Rice 1989). The estimator $\hat{\theta}$ of the parameter $F_{st}$ was calculated following Weir and Cockerham (1984). For electromorphs, we also estimated $\rho_{st}$, an analog of $F_{st}$ under the SMM (Rousset 1996), following Michalakis and Excoffier (1996). $F_{st}$ and $\rho_{st}$ were estimated both over all populations and between pairs of populations. Data analysis was conducted using the Genepop package, version 3.1d (Raymond and Rousset 1995).

Results

The SSCP method applied to electromorphs of the microsatellite locus BT1 revealed substantial additional variation within B. truncatus populations. Eight conformers were indeed detected for the three electromorphs considered. DNA sequencing confirmed that conformers differed only in the number of repeats of each repeated array of the compound core sequence (fig. 1). No nucleotide variation was detected in the 111-bp sequence of the flanking regions. Frequencies of electromorphs and conformers in each population are illustrated in figure 1. Size homoplasy was detected within four populations (Bouktra, Boundoum, Dyoro, and Tera D), resulting in an increase in gene diversity (table 1). The most striking example was observed in Tera D, where electromorph 184 corresponds to three conformers, increasing gene diversity from 0.04 for electromorphs to 0.62 for conformers. The proportion of heterozygotes observed for electromorphs was low in the populations studied and was not different from that estimated from conformers, since a single individual had only two different conformers (Dyoro population). As gene diversity increased with conformers and the frequencies of heterozygotes remained stable, the estimates of $F_{st}$ increased in populations in which size homoplasy was detected. An exception was Tera D, in which no heterozygote was detected either for electromorphs or conformers (table 1). This also holds for the estimates of the selfing rate. Exact tests for genotypic linkage disequilibria between BT1 and three other microsatellite loci (Viard, Justy, and Jarne 1997) were conducted on both conformers and electromorphs in the seven populations showing some electromorph variation (table 1). $P$ values were identical for electromorphs and conformers in Foua, Mari Nord, Canal, and Lambars, since the SSCP technique did not provide extra variability. In Boundoum and Dyoro, all $P$ values were higher than 0.10 for both electromorphs and conformers. On the other hand, they were higher than 0.10 for electromorphs and lower than 0.005 for conformers in Tera D. A single electromorph was fixed in Bouktra, so that the test for genotypic disequilibrium was not computed, while $P$ values were lower than 0.05 when conformers were considered.

Size homoplasy had striking consequences on the results of the analysis of population differentiation. For instance, several populations sharing the same electromorph (e.g., electromorph 180 in Fint and Foua; electromorph 184 in Bala, Lotzorai, and NPK) exhibited different conformers. This resulted in 43 of the 78 pairs of populations sharing no conformer, compared with only 16 for electromorphs. Only 19 pairs of populations remained unaffected by size homoplasy. Consequently, all exact tests of differentiation between pairs of populations conducted on conformers produced $P$ values lower than $10^{-4}$. On the other hand, the $P$ values were higher than 0.05 in 11 tests, between 0.05 and $10^{-4}$ in six tests, and lower than $10^{-4}$ in 55 tests when electromorphs were considered. Population differentiation tests could not be computed on six pairs of populations which were fixed for the same electromorph. However, those populations turned out to be highly differentiated when conformers were considered.

Pairwise $\hat{\theta}$ values computed from electromorphs and conformers are presented in figure 2. Note again
that they could not be computed on electromorphs for the six pairs of populations fixed for the same electromorph. When conformers were considered, those population pairs exhibited $\hat{\theta}$ values ranging between 0.172 and 1.000. $\hat{\theta}$ values were higher for conformers than for electromorphs for 31 pairs of populations, while the reverse was found for 22 pairs and identical values were obtained for 19 pairs. $\hat{\theta}$ values were higher for conformers than for electromorphs when populations shared identical electromorphs corresponding to different conformers. For instance, Lampsar and Bala were almost fixed for the same electromorph hiding different conformers (fig. 1): $\hat{\theta}$ increased from 0.069 to 0.906. On the other hand, the detection of size homoplasy for electromorphs not shared among populations resulted in a decrease in $\hat{\theta}$ values. For example, a decrease from 0.967 to 0.670 was observed for Arbatache and Tera D. Despite the change in many pairwise $\hat{\theta}$ values, $\hat{\theta}$ values computed over all populations using electromorph or conformer data were very similar (0.70 and 0.68, respectively). Finally, the $p_{st}$ value computed over all populations using electromorph variation was similar to the previous $\hat{\theta}$ values (0.75).

Discussion

The SSCP method has repeatedly been shown to be an efficient and versatile tool for population genetics purposes, as single-nucleotide variation can be detected without extensive sequencing (Bagley, Medrano, and Gall 1997; Angers and Bernatchez 1998). Orti, Pearse, and Avise (1997) showed that SSCP can also be used for analyzing homoplasmous variation at a microsatellite locus. However, these authors studied a pure microsatellite locus, so mutations were detected in the flanking regions only. We have shown here that SSCP can be used at the population level to uncover, at least partially, the size homoplasy occurring in the core sequence of compound microsatellite loci due to variation in the number of repeats of each repeated array. This means that more variation may be available at compound loci, as well as at both interrupted loci (unpublished data) and loci showing some variation in their flanking regions (Orti, Pearse, and Avise 1997), than when electromorphs are considered. Using SSCP also resulted in a significant improvement over previous studies considering the role of size homoplasy in population structure, such as those of Estoup et al. (1995) and Viard et al. (1998). These studies included a low number of sequenced electromorph copies, which limited their discriminating power, since the number of sequences detected per electromorph was found to increase with sequencing effort (Viard et al. 1998). In agreement with this result, the larger number of electromorph copies analyzed here allowed the detection of additional allelic forms within three of the populations previously studied by Viard et al. (1998). However, only a few extra sequences were detected using the SSCP method, suggesting that the relationship between the number of allelic forms detected and the sequencing effort may rapidly reach a plateau.

What are the consequences of size homoplasy on within-population structure? Previous studies suggested that size homoplasy reduces the number of alleles per population, the gene diversity, and the observed heterozygote frequency (e.g., Estoup et al. 1995; Garza and Freimer 1996; Angers and Bernatchez 1997; Viard et al. 1998; Taylor, Sanny, and Breden 1999). This was confirmed here in a population context. The SSCP method allows the uncovering of a concealed part of diversity, hence producing some more accurate estimates of intrapopulational diversity parameters. An interesting result is that heterozygote frequencies remained unchanged, while gene diversity increased considerably. As a consequence, estimates of $F_{st}$—and therefore of the selving rate—substantially increased in three populations. This result is unexpected, since the $F_{st}$-based method used here to infer the selving rate has been shown to be robust to mutation model and rate (Rousset 1996), so size homoplasy is not expected to downwardly bias estimates of selving rates in a population at mutation-drift-migration equilibrium. A possible situation that may lead to heterozygote deficiencies is the mixture of sampled individuals originating from differentiated subgroups (i.e., the “Walhund effect”). The ability to detect a Walhund effect is expected to increase with the level of polymorphism of the markers. This may explain why a larger amount of heterozygote deficiency was detected with conformers than with electromorphs. In agreement with this, significant genotypic linkage disequilibria were detected with conformers but not with electromorphs in Tera D and, to a lesser extent, in Bouktra.

Size homoplasy affects our perception of the distribution of variability among populations. Viard, Justy, and Jarne (1997) showed little heterogeneity in the distribution of BT1 electromorphs at a spatial scale lower than a few hundred kilometers. The present study clearly shows more heterogeneity when conformers are considered. This pattern is more similar to the one observed at more polymorphic microsatellite loci (Viard, Justy, and Jarne 1997). Additional variability detected by SSCP resulted in significant differentiation between all pairs of populations with conformers, while only 75% of tests were significant with electromorphs. A possible
reason is that higher levels of polymorphism increase the power of exact tests of population differentiation (Rousset and Raymond 1997). This was previously observed when microsatellite loci in *B. truncatus* (Viard, Justy, and Jarne 1997) were studied and when categories of markers with different levels of polymorphism were compared (e.g., microsatellites vs. allozymes, as in Estoup et al. 1998).

Quantification of population structure using *F*~ST~ indicated that uncovering size homoplasy may increase or decrease population differentiation, as both within- and among-population diversity influence estimates of population structure. The detection of size homoplasy does indeed increase the observed diversity among populations when populations share identical electromorphs corresponding to different conformers, while it increases within-population diversity when electromorphs not shared among populations hide several conformers. In any case, identity-based parameters of population structure estimated on conformers are likely to be more accurate than those calculated on electromorphs, as the number of allelic states is higher (Estoup and Angers 1998).

Size homoplasy leads to underestimation of genetic diversity within populations and differentiation among populations. Its effects are expected to be more marked for loci displaying low variability and for which a large fraction of electromorphs are shared among highly structured populations, as observed here. However, size homoplasy similarly affects diversity both within and among populations. As a consequence, *θ* values (or estimates of *p*~ST~) computed over all populations based on electromorphs and conformers are very similar. A general conclusion of our work is that detection of size homoplasy does not invariably produce higher estimates of population structure, contrary to what is sometimes asserted (e.g., Taylor, Sanny, and Breden 1999). This conclusion is consistent with the prediction, based on theoretical analyses, that size homoplasy in general should not lead to stronger population structure (Rousset 1996).

Two intriguing molecular results have to be highlighted because they may be relevant not only to the question of size homoplasy, but also to that of the mutation process at microsatellite loci. The first point bears on the variability of the microsatellite arrays studied. The BT1 locus has a GT~m~CT~m~ core sequence, with *n* varying from 9 to 13 and *m* varying from 2 to 5 for the conformers studied. That *n* varies is expected, given that pure repeats longer than about 10 units are generally variable (Weber 1990; reviewed in Estoup and Angers 1998). On the other hand, *m* shows a level of variation similar to that of *n* is surprising, given its low value. Variation in the number of repeats at short microsatellite stretches has been reported when orthologous loci have been compared across species (Messier, Li, and Stewart 1996; Primmer and Ellegren 1998; Taylor, Durkin, and Breden 1999), but not within species (Estoup et al. 1995), and even less within populations. Moreover, analyses conducted on large numbers of microsatellite loci extracted from DNA databases have indicated that dynamic mutation occurs for an array larger than four repeats (Jurka and Pethiyagoda 1995; Rose and Falush 1998). Mutation rates for very short stretches of repeats are probably very low. For instance, the average mutation rate for short microsatellites in *Drosophila* is about 5 × 10^{-6} (Schug, Mackay, and Aquadro 1997). A possible explanation for our results is that conformers represent very ancient lineages, but this seems unlikely given the absence of variation in the regions flanking the core sequence. Alternatively, the variability of consecutive arrays may be higher than if they were considered independently, as already suggested by Metzgar et al. (1998). The second point relates to the conformer diversity detected within some populations and between related populations. Size homoplasy was reported in four populations (fig. 1). The pattern observed in Tera D is puzzling, since there were three conformers for electromorph 184, and no additional electromorph. Genetic drift, which would have maintained electromorph 184 only, cannot be ruled out as an explanation for this result. However, electromorph 186 has not been reported in the vicinity of this population and is found only 200 km southeast of this area (Viard, Justy, and Jarne 1997). Moreover, electromorph 182 has never been reported in *B. truncatus*.

Altogether, these two points seem incompatible with a simple SMM, since this model assumes that mutation at locus BT1 should proceed through the addition/deletion of a single CT or GT repeat at a time. We propose two hypothetical mechanisms that may explain the pattern observed here and, more generally, some patterns of variation at compound dinucleotide loci. Let us assume a locus with a GT~CT~m~ core sequence, as in the present study. Under the first mechanism, a point mutation at the boundary between the GT stretch and the CT stretch may produce a new allele within the same electromorph. However, this requires a G↔C transversion mutation, which is known to occur at a low rate in most genomes. A similar mechanism based on the more frequent G↔A transition may explain homoplasious variation at other compound microsatellites (e.g., GT~AT~m~ loci). The second mechanism is based on the simultaneous occurrence (through slippage) of single-repeat-unit mutations in both arrays, for example, a deletion within the GT stretch and an addition within the CT stretch. Such a pattern has never been documented but is suggested by figure 2B in Levinson and Gutman (1987). Such a process would go undetected in pure loci. This second mechanism implies nonindependence in the mutation process of the two (or more) parts of a compound microsatellite (see also Metzgar et al. 1998). It is worth mentioning that this mechanism would result in a higher rate of size homoplasy than that under the standard SMM. Only intensive SSCP and/or sequencing programs could provide the empirical data needed to detail such a mechanism, since mutation would have to be looked for within electromorphs, and not among electromorphs, as is usually done.

**Acknowledgments**

We thank Patrice David and François Rousset for helpful discussion and comments, two anonymous re-
viewers for improving the manuscript, and M.-P. Dubois and C. Debain for technical help. This work was supported by a grant from the Centre National de la Recherche Scientifique (CNRS) to the Centre d’Ecologie Fonctionnelle et Evolutive (P.J.). B.A. benefited from a chercheur associé position with the CNRS.

LITERATURE CITED


PIERRE CAPY, reviewing editor

Accepted August 23, 2000