Genet. Res., Camb. (1998), **71**, *pp.* 213–222. *With* 1 *figure*. Printed in the United Kingdom © 1998 Cambridge University Press

The influence of mutation, selection and reproductive systems on microsatellite variability: a simulation approach

S. SAMADI[1]*, F. ERARD[2], A. ESTOUP[3] AND P. JARNE[1]

[1] *Génétique et Environnement – CC 065, Institut des Sciences de l'Evolution, Université Montpellier II, Place Bataillon, 34095 Montpellier cedex 5, France*

[2] *13 Bd Paul Vaillant-Couturier, 94800 Villejuif, France*

[3] *Laboratoire de Génétique des Poissons, CRJ-INRA, 78352 Jouy-en-Josas, France*

(*Received 15 October 1997 and in revised form 5 December 1997 and 17 February 1998*)

Summary

The influence of mutation, selection and reproductive systems on within-population size variation at microsatellite loci was analysed using simulations. Mutation occurred through either (biased or unbiased) replication slippage, or unequal recombination between homologous chromosomes. Selection acted either on large allele size, or on the difference in size between the two homologous alleles of an individual. Reproduction was either sexual (panmictic) or clonal. Classical population genetics parameters, such as gene diversity or variance of allele size, were followed over (generally) 5000 generations for various sets of values of the mutation rate and strength of selection, in either clonally or sexually reproducing populations. The reproduction system had little influence on genetic parameters, either under neutral conditions or when selection acted on large allele size. Selection against difference in allele size strongly constrained variability in panmictic populations, whereas a limited influence was observed in clonal populations. Selection against the difference in allele size between the two alleles of an individual is an alternative explanation for the long life expectancy of microsatellite loci in sexual species. Whether this selection process actually occurs can therefore be tested by comparing the allele size distribution of microsatellite loci between regions/genomes exhibiting markedly different recombination rates.

1. Introduction

Theoretical analyses of microsatellite variability have generally been conducted at equilibrium between genetic drift and mutation in populations of constant size (review in Freimer & Slatkin, 1996; Jarne & Lagoda, 1996). Several theoretical models of mutation have been considered, including the stepwise mutation model (SMM; Ohta & Kimura, 1973), its offspring the two-phase model (TPM; DiRienzo *et al.*, 1994) and the infinite alleles model (IAM; Crow & Kimura, 1964). The expected variability can also be derived in more complex situations, such as when the population size is variable (DiRienzo *et al.*, 1994; Cornuet & Luikart, 1996). However, the mutational mechanisms of microsatellite loci remain a matter of debate. Two molecular processes generating variability have been proposed: replication slippage due to slipped-strand mispairing and unequal recombination between homologous chromosomes (or chromatids). Empirical studies support the first process (Levinson & Gutman, 1987; Schlötterer & Tautz, 1992; Strand *et al.*, 1993). This issue has also been addressed using computer simulations. For example. Stephan & Cho (1994) showed that unequal exchanges between sister chromatids play a marginal role in microsatellite evolution.

The models mentioned above do not consider some parameters relevant to the evolution of microsatellite variability. First, analysis of collections of mutants suggests that mutation might be a (upwardly) biased process (Amos *et al.*, 1996; Primmer *et al.*, 1996). Primmer *et al.* (1996) also showed that larger alleles may be more prone to mutation, and Amos *et al.* (1996) that mutation is more likely to occur in

* Corresponding author. Phone: (+33) 04 67 14 46 29. Fax: (+33) 04 67 14 36 22. e-mail: samadi@isem.univ-montp2.fr.

individuals with a large difference in allele size. Biased mutation was considered by Garza *et al.* (1995) and Zhivotovski *et al.* (1997) in a model in which large (small) alleles preferentially mutate towards smaller (larger) alleles. The theoretical expectations of population variability derived under this model fit the data better than with unbiased mutation (Garza *et al.*, 1995).

Secondly, various forms of selection may act on the evolution of microsatellites (reviewed in Garza *et al.*, 1995). Microsatellites may play a general role in chromosome pairing during meiosis, with no specification of the action of selection at each locus (Pardue *et al.*, 1987). This was suggested by the uniform distribution of dinucleotide microsatellites along chromosomes. Selection may also act as a truncating mechanism, imposing a ceiling on allele size. Alleles indeed have a finite size, generally shorter than a few tens of repeat units, and large alleles at microsatellite loci associated with genetic diseases are strongly counter-selected (Sutherlands & Richards, 1995). We further suggest that selection may act on the difference in size between the two alleles borne by homologous chromosomes in a diploid individual. This may derive from microsatellites ensuring a correct alignment of homologous chromosomes at meiosis (Pardue *et al.*, 1987). Too large a difference would destabilize chromosome pairing, and provoke abnormal recombination in the regions flanking microsatellites. This hypothesis implies selection against heterozygosity as a function of the difference in size between homologous alleles. A similar idea has been proposed, although in a different context (Stephan & Langley, 1992). These authors suggested that the efficiency with which double-stand breaks in DNA are repaired depends on the heterozygosity surrounding the breaks. This process strongly depresses the variability maintained within populations.

Thirdly, the reproductive system may also be a relevant factor, although one that has received little attention. For example, meiosis does not occur in some organisms (review in Bell, 1982). This makes little difference for microsatellite variability if mutation occurs only through mitotic replication slippage, though not if it also occurs through unequal exchange during meiosis. Moreover, selection acting on size difference between alleles borne by a given individual is expected to act differentially in asexual and sexual species. In the latter, any allele at a given locus may indeed be paired with any other allele of the population through reproduction, whereas pairs of alleles are maintained within clonal lineages.

These parameters have also to be considered when analysing the life expectancy of microsatellite loci. Homologous polymorphic loci are very often found between related species in cross-priming studies (see, for example, FitzSimmons *et al.*, 1995), suggesting that their fixation into low mutating states (*i.e.* with fewer than three repeats) is prevented. Tachida & Iizuka (1992) suggested that a mutation bias towards the addition of repeats would both prevent the fixation of microsatellite loci to low mutating states and facilitate the emergence of new loci. We here suggest that selection against the difference in size between the two alleles of an individual could stabilize allelic distributions. This should indeed maintain allelic size within a given range, and hence prevent the fixation into low mutating states.

We developed a simulation model in order to explore the influence on microsatellite variability and on the life expectancy of loci of the processes mentioned above, namely, mutation (replication slippage or unequal exchange between homologous chromosomes), reproductive systems (clonal or panmictic reproduction) and selection (acting on absolute allele size or on size difference between the two alleles at a locus on homologous chromosomes). Comparing the results of our model with those of previous models (e.g. DiRienzo *et al.*, 1994; Zhivotovski *et al.*, 1997) proved to be difficult. However, our model allowed investigation of the effects of associations between several parameters of biological relevance (such as various mutation models, selection or mating systems) that are not easy to handle analytically.

2. Materials and methods

(i) *The model*

The source program of the model, written in C, is available upon request from S. Samadi. The acronyms used in the text, variables of the model and parameters describing the allelic distributions are defined in Table 1. The model simulated the evolution of a microsatellite locus in a diploid population of constant size N. Each generation, the population was characterized by a genotypic matrix M_G of terms m_{ij} representing the number of individuals with alleles of size (number of repeats) i and j ($i \leqslant j$). Individuals underwent, every generation, selection and mutation, and then reproduced. They initially had genotype (L, L), so that $m_{LL} = N$, and $m_{ij} = 0$ for $i \neq L$ or $j \neq L$.

Selection took two forms, both acting at the diploid stage. The first acted on the maximal size T gained by alleles. An allele of size i had an associated fitness $w_i = 1$ for $i \leqslant T$, and $w_i = \exp(2 \ln 10(T-i)/T)$ for $i > T$. This arbitrarily chosen function ensured a decrease in fitness with increasing allelic size. Individual fitness was calculated as $W_{ij} = w_i w_j$. The second type of selection played on the difference in size between the two alleles of a given individual. Large differences may indeed disturb meiosis, because of mispairing of homologous chromosomes (see Section 1). Selection acted when the difference was larger than the selection

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

Table 1. *Acronyms used in the text, variables of the model and parameters describing the allelic distributions*

Acronyms	
SSM	Slipped-strand mispairing
SMM/TPM	Stepwise mutation model/two-phase model
IAM	Infinite alleles model
UEHC	Unequal exchange between homologous chromosomes
Variables	
N	Number of individuals
L	Initial length of alleles
Mit	Number of mitoses per individual
S	Rate of slippage per mitosis
u	Mutation rate through replication slippage
B	Slippage bias towards deletion
R	Recombination threshold
rec	Rate of UEHC per generation
u'	Total rate of mutation (replication slippage and UEHC)
T	Selection threshold
Parameters	
m	Mean size (in number of repeats)
Var	Variance
H_e	Gene diversity
RS	Range size
n	Number of alleles
M	Number of modes
A	State of the class 'one repeat'

threshold T. Fitness of genotypes $(i,j$ with $i<j$ was given by: $w_{ij}=1$ for $j-i\leqslant T$, and $w_{ij}=\exp(2\ln 10(T-(j-i))/T)$ for $j-i>T$. This function was also arbitrarily chosen such that fitness decreased with increasing difference in size of homologous alleles.

Mutation occurred at the diploid stage. The model mimicked a germ line undergoing mitosis and recombination occurring during the first meiotic division. Alleles evolved under either replication slippage alone, or replication slippage and unequal exchange between homologous chromosomes (UEHC). Alleles of size one repeat did not mutate and this allelic class was thus an absorbing state. Replication slippage inserted or deleted one repeat unit at mitosis. Diploid individuals were subjected to Mit successive mitoses, producing 2^{Mit} cells from an initial cell. Two cells were produced after each mitosis, one presenting the parental alleles while the other differed at one of the two parental alleles by one repeat with a probability S ($S/2$ for deletion and insertion). The mutation rate per allele and mitosis was $S/2$ ($S/4$ for deletion and $S/4$ for insertion). The total mutation rate per generation was $u=1-(1-S/4)^{Mit}\approx(Mit\,S)/4$. With large values of S and Mit, some alleles are expected to mutate more than once per generation. However, the frequency of such events was negligible compared with single-step mutation (the ratio of two-step over single-step events was indeed of the order $S(Mit-1)/8$, that is around S given the values used in our simulations). Overall this procedure fitted the SMM. We also introduced a mutational bias, such that alleles could increase or decrease in number of repeats with probabilities $(1-B)S$ and BS respectively. $B=\frac{1}{2}$ corresponds to the situation described above. After Mit mitoses, the frequency of each genotype (i,j) was calculated. In the simulations with both slippage and UEHC, UEHC was initiated only when both alleles from a given individual were longer than a threshold R. Homologous recombination indeed seems to depend on a minimum perfect homology length (Fujitani *et al.*, 1995). Among individuals with $(i,j)\geqslant(R,R)$, a proportion rec mutated. For each individual, a number α lying between $(R-\min(i,j))$ and $(\max(i,j)-R)$ was randomly chosen. As this procedure involved random drawing among individuals, the populations studied here are nearly infinite (in other words, we did not work on frequencies here). $|\alpha|$ was the size of the fragment exchanged between alleles. The genotype after recombination was (i',j') with $i'=i+\alpha$ and $j'=j-\alpha$. The minimum size of the exchanged fragment was equal to the size of the threshold R.

Reproduction was either panmictic or clonal. In the panmictic procedure, the allelic frequencies were calculated in the gametic population after mutation, and genotypes were then reconstituted in proportion to these frequencies: $M_{ii}=f_i^2N$, and $m_{ij}=2f_if_jN$ for $i\neq j$, with f_i and f_j the frequencies of alleles of length i and j respectively. In the clonal procedure, the population was reduced from $2^{mit}N$ diploid gametes (after Mit mitoses with no meiotic reduction) to N individuals proportionally to the allelic frequencies in the gametic population.

(ii) *Simulations and parameters estimated*

Simulations were run to investigate the influence on within-population variability of a mixed mode of mutation (slippage, or slippage and UEHC), reproduction (clonal or panmictic) and selection (on large allele size and difference in allele size). This allowed investigation of various forms of reproduction, including apomictic and automictic parthenogenesis (see Bell, 1982, chapter 1). We simulated a population of size $N=1000$ individuals, but it can be considered as nearly infinite since frequencies were used. As the equilibrium in our model is the one repeat allelic class, we chose (after initial trials) to follow populations over 5000 generations. However, some simulations were run over 50000 generations to analyse the influence of the number of generations and the long-term persistence of alleles. The values of the variables L, B, R, rec and T were defined at the beginning of each run. The number of mitoses Mit

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

Table 2. *Microsatellite variability in panmictic populations*

Variables					Parameters									
												A		
L	*B*	*rec*	*R*	*u′*[a]	H_e	*m*	*Var*	*RS*	*n*	*M*	F_1	0	1	2
10	0·5	—	—	—	0·91 (0·01)	9·78 (0·38)	11·93 (2·32)	15·88 (0·97)	15·44 (0·89)	3·12 (0·46)	0·03 (0·01)	16	8	1
20	0·5	—	—	—	0·90 (0·01)	19·42 (0·33)	9·75 (1·60)	14·90 (1·05)	14·80 (0·99)	2·73 (0·43)	0·00	25	0	0
20[b]	0·5	—	—	—	0·91 (0·01)	19·49 (0·48)	11·21 (1·87)	17·00 (1·24)	16·80 (1·30)	3·20 (0·91)	0·00	10	0	0
10	0·45	—	—	—	0·90 (0·01)	10·55 (0·33)	9·21 (1·65)	15·13 (1·13)	14·87 (1·08)	2·73 (0·65)	0·00	15	0	0
10	0·35	—	—	—	0·90 (0·01)	12·53 (0·26)	10·04 (2·12)	15·13 (1·42)	14·87 (1·19)	2·88 (0·74)	0·003 (0·006)	14	1	0
10	0·5	10^{-2}	15	2·84 (0·37)	0·91 (0·01)	10·86 (0·75)	11·66 (4·17)	16·40 (2·95)	16·00 (2·40)	4·00 (1·24)	0·00	5	0	0
10	0·5	10^{-3}	15	2·51 (0·003)	0·93 (0·00)	8·78 (0·39)	17·56 (3·73)	17·80 (1·14)	16·80 (1·30)	5·00 (1·24)	0·05 (0.03)	1	4	0
20	0·5	10^{-2}	15	7·61 (0·59)	0·94 (0·02)	18·25 (1·46)	35·74 (18·22)	33·00 (6·53)	27·20 (4·18)	8·40 (2·81)	0·00	5	0	0
20	0·5	10^{-3}	15	3·08 (0·11)	0·94 (0·01)	18·74 (1·49)	35·66 (9·49)	26·67 (3·23)	24·33 (2·77)	7·73 (1·20)	0·00	15	0	0
20	0·5	10^{-2}	20	4·96 (0·78)	0·90 (0·01)	19·59 (0·87)	10·96 (4·07)	17·80 (2·59)	17·60 (2·74)	3·60 (1·59)	0·00	5	0	0

Means (and 95% confidence intervals) for parameters describing the genetic variability in simulations with panmictic populations evolving under either unbiased replication slippage (upper part), biased replication slippage (middle part) or unbiased replication slippage and UEHC (lower part). Simulations were run over 5000 generations. *L*, *B*, *rec*, *R* and *u′* are the initial size, the replication bias, the fraction of individuals recombining, the recombination threshold and the actual mutation rate respectively.

H_e, *m* and *Var*, *RS*, *n* and *M* are the gene diversity, the mean and variance of allele size, the range size of alleles, the number of alleles and the number of modes respectively. F_1 is the frequency of the one repeat allelic class. A indicates the number of simulations for which the one repeat allelic class is not a mode ($A = 0$), is a mode ($A = 1$) or is the major mode ($A = 2$) of the distribution.

[a] *u′* is to be multiplied by 10^{-3}. $u = 2{\cdot}5 \times 10^{-3}$ except [b] for which $u = 3{\cdot}0 \times 10^{-3}$. —, irrelevant.

was 10 or 20, and the rate of mutation per mitosis *S* was between 10^{-4} and 10^{-2}; *u* therefore lay between $2{\cdot}5 \times 10^{-4}$ and $2{\cdot}5 \times 10^{-2}$, corresponding to values reported in the literature (see Amos *et al.*, 1996; Primmer *et al.*, 1996). *rec* was chosen such that mutation through recombination remained of limited magnitude when compared with slippage. With both replication slippage and UEHC, the actual mutation rate (*u′*) was calculated as the sum of the mutation rate through replication slippage and the mutation rate through UEHC. This latter differed from *rec* because it concerned only those individuals harbouring two alleles of size over the recombination threshold (*R*).

The allelic distributions were described using the following classical parameters measured at the end of each run: gene diversity (H_e), mean (*m*) and variance (*Var*) of allele size, range size of alleles (*RS*), number of alleles (*n*) and of modes (*M*) as defined in Shriver *et al.* (1993). In order to check for the absorption of allelic distributions into the one repeat allelic class, we defined a three-state variable (*A*). $A = 0$ when the one repeat allelic class is not a mode of the distribution, $A = 1$ when this class is a mode of the distribution and $A = 2$ when this class is the major mode of the distribution. The frequency of the one repeat allelic class was also followed.

3. Results

We first considered panmictic populations evolving under replication slippage only with no selection (hereafter referred to as 'the basic situation'), with $u = 2{\cdot}5 \times 10^{-3}$ and $L = 20$. Results are given in Table 2. Modifying *L* only translates, as expected, allelic distributions towards smaller or larger size (Table 2 for $L = 10$ and 20; results not shown for $L = 25$). Distributions are more frequently absorbed into the one repeat class for low *L* and high *u* (Table 2). We thereafter generally used conditions avoiding absorption into this class. Introducing a bias in the mutation process mainly translated the distributions towards larger alleles, but other genetic parameters remained essentially unchanged (Table 2). The influence

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

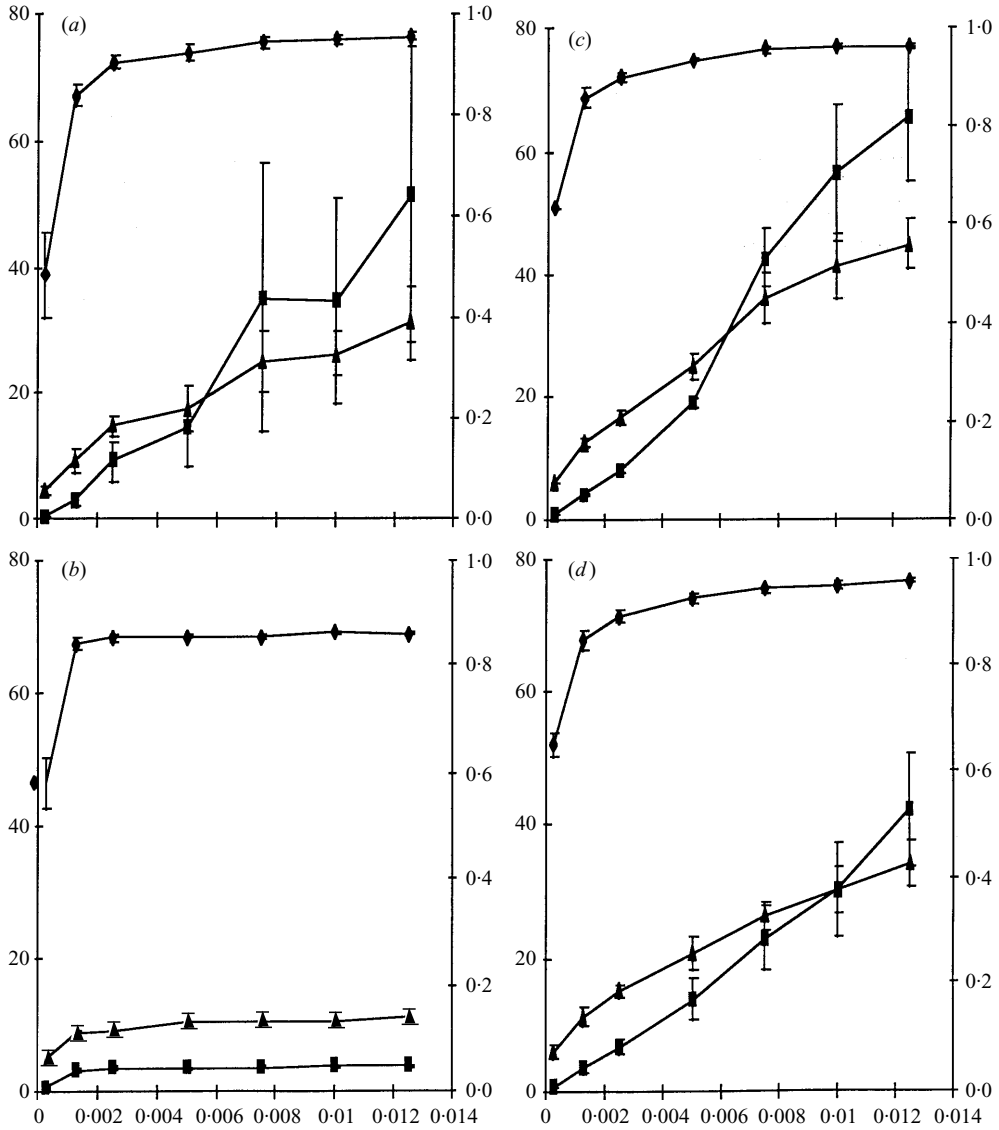

Fig. 1. Gene diversity (H_e, on right-hand ordinate; diamonds), variance in allele size (*Var*, on left-hand ordinate; squares) and number of alleles (n on left-hand ordinate; triangles) for various mutation rates u (abscissa) in populations evolving under unbiased replication slippage only. Vertical bars correspond to 95% confidence intervals. The initial size of the allele, L, was 20 repeats. (*a*) Panmictic reproduction and no selection. (*b*) Panmictic reproduction and selection against the difference in allele size. The selection threshold, T, was seven repeats. (*c*) Clonal reproduction and no selection. (*d*) Clonal reproduction and selection against difference in size with $T = 7$.

of the mutation rate was analysed running simulations with u varying from $2{\cdot}5 \times 10^{-4}$ to $2{\cdot}5 \times 10^{-2}$. Gene diversity quickly reached a plateau, while both the allelic size variance and the number of alleles increased roughly linearly (Fig. 1*a*).

We explored the role of UEHC in panmictic populations, simulating a mixed model of mutation (slippage and UEHC) (Table 2). For a given u, this had little influence on the mean size of alleles, though produced much wider distributions, as indicated by the larger variance of allele size. This did not result only from a higher overall mutation rate than in the basic situation, since simulations with similar overall mutation rates showed the same tendency (compare lines 3 (slippage only) and 9 (slippage and UEHC) in Table 2). Consequently, distributions generated with UEHC were more easily absorbed into the one repeat class, especially with low initial size of alleles. When initial size of allele was higher, A was equal to zero but the slight decrease in mean allele size indicated that the absorption of the distribution into the one repeat allelic state began.

The above situations were analysed in clonally reproducing populations, and compared with panmictic populations. With both unbiased and biased slippage, as well as with UEHC, narrower distributions (lower variance and number of modes) were obtained in clonal population for both $L = 10$ and $L = 20$, but the other parameters estimated remained very similar (compare Tables 2 and 3). The curves

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

Table 3. *Microsatellite variability in clonal populations*

Variables					Parameters									
												A		
L	*B*	*rec*	*R*	*u′*[a]	H_e	*m*	*Var*	*RS*	*n*	*M*	F_1	0	1	2
10	0·5	—	—	—	0·87 (0·02)	10·36 (0·37)	5·07 (0·96)	14·00 (1·07)	14·00 (0·73)	1·80	0·00	5	0	0
20	0·5	—	—	—	0·89 (0·01)	20·04 (0·14)	6·40 (0·42)	14·45 (0·56)	14·45 (0·56)	1·65 (0·29)	0·00	25	0	0
20[b]	0·5	—	—	—	0·91 (0·00)	19·97 (0·23)	9·21 (0·75)	17·20 (0·96)	17·20 (0·96)	1·90 (0·62)	0·00	10	0	0
10	0·45	—	—	—	0·89 (0·01)	10·63 (0·21)	6·64 (1·31)	14·40 (1·71)	14·40 (1·71)	1·80 (0·73)	0·00	5	0	0
10	0·35	—	—	—	0·89 (0·01)	12·03 (0·40)	6·76 (1·34)	15·25 (1·73)	15·25 (1·73)	1·75 (0·32)	0·00	8	0	0
10	0·5	10^{-2}	15	2·51 (0·03)	0·91 (0·01)	9·79 (0·42)	9·89 (3·51)	16·80 (1·14)	16·80 (1·14)	2·80 (1·14)	0·01 (0·02)	4	1	0
20	0·5	10^{-2}	15	9·17 (0·49)	0·95 (0·00)	20·30 (0·59)	34·76 (5·02)	34·00 (2·48)	33·60 (2·74)	6·20 (0·73)	0·00	5	0	0
20	0·5	10^{-3}	15	3·19 (0·04)	0·95 (0·00)	19·76 (0·29)	26·61 (1·79)	30·00 (1·39)	29·33 (1·41)	5·13 (0·71)	0·00	15	0	0
20	0·5	10^{-2}	20	5·41 (0·59)	0·92 (0·00)	20·11 (0·48)	11·52 (0·18)	22·00 (1·07)	21·80 (1·14)	1·80 (0·73)	0·00	5	0	0

Means (and 95% confidence intervals) for parameters describing the genetic variability in simulations with clonal populations evolving under either unbiased replication slippage (upper part), biased replication slippage (middle part) or unbiased replication slippage and UEHC (lower part). Simulations were run over 5000 generations. Variables and parameters as in Table 2.

[a] u' is to be multiplied by 10^{-3}. $u = 2{\cdot}5 \times 10^{-3}$ except [b] for which $u = 3{\cdot}0 \times 10^{-3}$. —, irrelevant.

Table 4. *Microsatellite variability after 50000 generations*

		Variables		Parameters									
											A		
Population	Selection	*B*	*T*	H_e	*m*	*Var*	*RS*	*n*	*M*	F_1	0	1	2
P	NS	0·5	—	0·20 (0·22)	1·5 (0·53)	2·15 (2·09)	8·6 (0·89)	6·2 (3·42)	2·8 (0·84)	0·88 (0·13)	0	0	5
P	S1	0·5	20	0·44 (0·26)	2·25 (1·15)	5·08 (5·17)	9·00 (2·55)	8·80 (2·77)	3·60 (1·34)	0·72 (0·19)	0	0	5
P	S1	0·35	20	0·91 (0·03)	9·21 (1·59)	20·65 (8·42)	17·6 (2·61)	17·60 (2·61)	4·6 (1·52)	0·09 (0·11)	2	1	2
P	S2	0·5	7	0·84 (0·01)	21·24 (1·83)	3·11 (0·36)	9·40 (0·55)	9·40 (0·55)	1·20 (0·45)	0·00	5	0	0
C	NS	0·5	—	0·96 (0·01)	24·29 (2·50)	66·55 (25·63)	38·8 (5·72)	38·20 (5·81)	9·40 (1·52)	0·00 (0·01)	5	1	0
C	S1	0·5	20	0·94 (0·00)	12·23 (0·61)	20·32 (1·57)	20·60 (0·55)	20·60 (0·55)	4·40 (1·14)	0·02 (0·02)	2	3	0
C	S1	0·35	20	0·90 (0·01)	15·42 (0·38)	8·98 (1·42)	15·8 (0·84)	15·8 (0·84)	3·80 (0·84)	0·00	5	0	0
C	S2	0·5	7	0·95 (0·00)	24·35 (2·39)	31·17 (6·03)	26·60 (2·70)	26·60 (2·70)	5·80 (0·84)	0·00	5	0	0

Mean values (and 95% confidence intervals) for parameters describing the genetic variability in simulations with panmictic (P) or clonal (C) populations evolving under either unbiased or biased replication slippage. NS is the neutral situation. Selection is either against large allele size (S1) or against the difference in size between alleles (S2). Simulations were run over 50000 generations. Variables and parameters as in Table 2. T is the selection threshold. $u = 2{\cdot}5 \times 10^{-3}$, $L = 25$. —, irrelevant.

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

obtained with variable mutation rates were similar in shape to those obtained in panmictic populations, although the variance of allele size and number of alleles were much lower in clonal populations (Fig. 1*c*). Although the absorption of alleles into the one repeat allelic state was less important for clonal populations than for panmictic populations, the same rules were followed (for example, absorption is increased with the mixed mode of mutation).

The influence of selection against size difference was analysed under slippage, D and L taking values of 7 and 25 respectively, so as to avoid absorption into the one repeat class. In panmictic populations selection little affected gene diversity, but dramatically reduced the variance in allele size, and the number of both alleles and modes (Fig. 1*b*), producing 'bell-shaped' distributions (results not shown). The effect was much weaker in clonal populations, although the genetic parameters never reached the values obtained without selection (Fig. 1*d*). All curves in Fig. 1*b* reached a plateau over a mutation rate of about 10^{-3}, while a similar trend was not observed without selection.

Simulations over 50000 generations were run under replication slippage only. We followed both panmictic and clonal populations, without or with selection (Table 4). The one repeat class was, as expected, more often preponderant than previously, especially in panmictic populations, which makes comparisons across situations pointless. Both biased slippage associated with selection against large allele size and selection against size difference slowed down the absorption process. However, the selection against difference in allele size was more efficient in maintaining the allele size distribution. The important tendency observed over 5000 generations (Fig. 1) was confirmed though: selection against large allele size lowered the variability in all situations, while selection against size difference was much more efficient for panmictic than for clonal reproduction.

4. Discussion

The basic situation analysed corresponds to that considered by previous authors for analysing microsatellite variability at equilibrium between mutation (SMM) and genetic drift, without (Ohta & Kimura, 1973; Valdès *et al*., 1993) or with (Feldman *et al*., 1997) constraints on allele size. However, our results cannot be compared directly with those from analytic formulas derived by these authors, since genetic drift was not considered in the present study. A peculiarity of our model is that the one repeat allelic class is an absorbing state, meaning that there is a lower ceiling to allele size. With no mechanism avoiding absorption into this allelic class, this latter constitutes the equilibrium state of our simulations. Such a constraint is taken into account in only a few models, although in a slightly different form (Nauta & Weissing, 1996; Feldman *et al*., 1997). Its biological meaning is to be related to the probably very low rate of mutation of very short alleles (Jin *et al*., 1997). This class reached particularly high frequencies when the mutation rate was high. Of course, its frequency was lower when an upward bias in the slippage process was introduced. Our simulations also show that this bias had little influence on the parameters measured, and mainly shifted distributions towards larger alleles. However, a bias was required when the initial size was low, or the mutation rate high, to avoid absorption of allelic distributions into the one repeat class. More generally, such a bias facilitates the emergence of microsatellite loci from very short repeat arrays (Tachida & Iizuka, 1992), and empirical data support the idea that mutation is upwardly biased (Amos *et al*., 1996; Primmer *et al*., 1996). Simulations over 50000 generations confirm that biased mutation towards large allele size slows down the fixation of loci into the one repeat allelic class. Mutation including both slippage and UEHC produced much wider distributions than with slippage only for similar overall mutation rate. UEHC indeed generates variants differing by more than one repeat from the progenitor allele. This is formally equivalent to a comparison between the SMM and the TPM, the second producing a much larger variance of allele size (DiRienzo *et al*., 1994). It can certainly be argued that there is no evidence for the role of UEHC in the evolution of most microsatellite loci. This was the very reason for giving much more weight in our simulations of a mixed process of mutation to slippage than to UEHC. Moreover, the mutation model may vary with the loci considered within the same genome, as suggested by results in man (DiRienzo *et al*., 1994) and bees (Estoup *et al*., 1995).

Two selection procedures were considered. The first acted on the maximum size of alleles. This was suggested by the finite size of alleles at microsatellite loci (generally below 100 repeats), but also by the relationship between the number of repeats with both the molecular activities of particular genes including microsatellite sequences (see Kashi *et al*., 1997) and some neurodegenerative diseases (Sutherlands & Richards, 1995). However, we were concerned with the bulk of loci that are located outside genes. Our simulations showed that selection against large size (associated with biased mutation) similarly decreased within-population variability in both panmictic and clonal populations. Selection acting on the difference in allele size also reduced variability. Stephan & Langley (1992) obtained a similar result, using a model in which local heterozygosity reduces the efficiency of DNA mismatch repairs after double-strand breaks. However, these authors considered neither repetitive sequences (they analysed nucleotide

https://doi.org/10.1017/S0016672398003292 Published online by Cambridge University Press

sequences evolving under point mutation) nor the influence of reproduction. An important result of our simulations is that the reduction in variability was much more severe in panmictic than in clonal populations when selection acted on the difference in allele size, though not with selection on absolute size. In panmictic populations, alleles of very different size may indeed be paired each generation, while such a situation results from mutation alone in clonal populations.

Allelic distributions are currently available in too few clonal species for a comparison between reproductive systems. An alternative is to compare size distributions of cloned loci that have been obtained using similar techniques. Assuming that the loci cloned in a given species follow the same mutational law and are genetically independent, distributions based on a single cloned allele per locus are equivalent to distributions of alleles at a single locus. The number of cloned loci in the former case, and the number of alleles per locus in the latter, should also be large enough to ensure that all, or nearly all, possible allelic states are represented. As mentioned above, this value is likely to be constrained (see Nauta & Weissing, 1996; Feldman *et al*., 1997). Such an analysis performed in two insect species (Estoup *et al*., 1993), two fish species (unpublished data) and humans (Weber, 1990) shows that distributions in sexual species typically have a small mean, and are more or less 'bell-shaped'. On the other hand, a more uniform distribution with a very large variance was found in the parthenogenetic freshwater snail *Melanoides tuberculata*, suggesting that selection acts on the difference in allele size at microsatellite loci (Samadi *et al*., 1998).

However, this test is associated with two pitfalls. First, selection may indirectly act on microsatellites through, for example, background selection (Charlesworth *et al*., 1993). This process strongly depresses the within-population variability when recombination is reduced (e.g. in selfing or clonal species or on the Y chromosome), whatever the mutation process (Slatkin, 1995; Viard *et al*., 1997). It seems difficult to predict its influence. Secondly, we must assume that the distribution of life expectancies across loci has a moderate variance. Whatever the value of this test, our results suggest that it may be important to consider the interaction between the reproductive system and selection when analysing the evolution of microsatellites. More generally, this should be kept in mind when contrasting the variability of genomes, or regions of genomes, with various rates of recombination.

Our simulations also suggested that allelic distributions are less easily absorbed into the one repeat class in the long term (over 50000 generations) when selection acts on size difference than in the other situations simulated, especially when selection was on size. Selection on size difference indeed maintains the cohesion of allelic distributions, preventing absorption. This mechanism is particularly efficient at maintaining allelic distributions within a particular range size of alleles, and then avoiding the absorption into a low mutating state. This further suggests that mutation and genetic drift alone cannot explain the evolution of microsatellite loci in the long term. One has to assume that allele size has both upper and lower limits (Garza *et al*., 1995). This may result from selection, as mentioned above. An alternative is an asymmetric mutation process, such as a bias towards large alleles (see above). However, neither stabilizing (Garza *et al*., 1995; Zhivotovski *et al*., 1997) nor biased mutation should produce any difference between sexual and clonal species. On the other hand, a difference is expected if mutation depends on size difference at the individual level (Amos *et al*., 1996). However, whether this will in general differentially affect sexual and clonal species is unclear. Their relative heterozygosity indeed depends on the origin of clonality. For example, heterozygosity may be high when clonality is associated with polyploidy. It is worth mentioning that selection against the difference in size between the two alleles of an individual can explain both the ceiling on allele size and the long life expectancy of microsatellites in panmictic populations. This mechanism is more efficient at constraining allelic size distributions within a particular range than the alternative hypothesis (e.g. association of upwardly biased mutation and upper allele size constraint; Table 4). However, the relative efficiency of these two processes depends on the parameter values.

Alleles in our model were defined by their number of repeats, meaning that we did not distinguish between types (e.g. di- or tetra-nucleotides) of loci. However, for a given number of repeats, a tetranucleotide locus is longer than a di-nucleotide. This may affect the selective pressure, if the stability of meiotic processes depends on the absolute size (in base pairs) of microsatellites. Loci with a larger repeat unit should therefore experience stronger selection (against the difference in size), especially in genomic regions with high recombination rates. Tetra-nucleotides, when compared with di-nucleotides, should be less variable in regions of high recombination. Tetranucleotide loci are indeed less polymorphic than di-nucleotides in the human genome (Chakraborty *et al*., 1997), but the generality of this result has to be investigated (see Viard *et al*., 1996 for a different result). Moreover, the genomic positions of these loci were not known. Another prediction is that tetranucleotide loci should have both a lower density and a more heterogeneous distribution in the genome. On the other hand, they should be as frequent and variables di-nucleotides in non-recombining regions. Di-nucleotides are indeed more frequent and evenly

distributed than tetra-nucleotides (references in Jarne & Lagoda, 1996). However, the distribution of microsatellites in connection with recombination rates is poorly known. Finally, tri-nucleotides are very often associated with genes, and are likely to be under other selective pressures.

The authors are indebted to P. David, F. Viard and three anonymous referees for discussions and comments on the manuscript, and to K. Gharbi for access to unpublished cloning data. This work was supported by a grant from GREG (94/84). S.S. benefited from a fellowship from the Ministrère de l'Enseignement Supérieur. This is contribution 98.058 of Institut des Sciences de l'Evolution.

References

Amos, W., Sawcer, S. J., Feakes, R. W. & Rubinzstein, D. C. (1996). Microsatellites show directional bias and heterozygote instability. *Nature Genetics* **13**, 390–391.

Bell, G. (1982). *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. Berkeley: University of California Press.

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J. & Deka, R. (1997). Relative mutation rates at di-, tri- and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences of the USA* **94**, 1041–1046.

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

Cornuet, J.-M. & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Crow, J. F. & Kimura, M. (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

DiRienzo, A., Peterson, A. C., Garza, J. C., Valdès, A. M., Slatkin, M. & Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA* **91**, 3166–2170.

Estoup, A., Solignac, M., Harry, M. & Cornuet, J.-M. (1993). Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Research* **21**, 1427–1421.

Estoup, A., Tailliez, C., Cornuet, J.-M. & Solignac, M. (1995). Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution* **12**, 1074–1084.

Feldman, M. W., Bergman, A., Pollock, D. D. & Goldstein, D. B. (1997). Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* **145**, 207–216.

FitzSimmons, N. N., Moritz, C. & Moore, S. S. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* **12**, 432–440.

Freimer, N. B. & Slatkin, M. (1996). Microsatellites: evolution and mutational processes. In *Variation in the Human Genome* (ed. D. Chadwick & G. Cardew), pp. 51–72. Chichester: Wiley.

Fujitani, Y., Yamamoto, K. & Kobayashi, I. (1995). Dependence of frequency of homologous recombination on the homology length. *Genetics* **140**, 797–809.

Garza, J. C., Slatkin, M. & Freimer, N. B. (1995). Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Molecular Biology and Evolution* **12**, 594–603.

Jarne, P. & Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424–429.

Jin, L., Macaubas, C., Hallmayer, J., Kimura, A. & Mignot, E. (1997). Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proceedings of the National Academy of Sciences of the USA* **93**, 15285–15288.

Kashi, Y., King, D. & Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* **13**, 74–78.

Levinson, G. & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**, 203–221.

Nauta, M. J. & Weissing, F. J. (1996). Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**, 1021–1032.

Ohta, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research Cambridge* **22**, 201–204.

Pardue, M. L., Lowenhaupt, K., Rich, A. & Nordheim, A. (1987). (dC-dA)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO Journal* **6**, 1781–1789.

Primmer, C. R., Ellegren, H., Saino, N. & Moller, A. P. (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics* **13**, 391–393.

Samadi, S., Artiguebielle, E., Estoup, A., Pointier, J.-P., Silvain, J.-F., Heller, J., Cariou, M.-L. & Jarne, P. (1998). Density and variability of dinucleotide microsatellites in the parthenogenetic polyploid snail *Melanoides tuberculata*. *Molecular Ecology*, in press.

Schlötterer, C. & Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* **20**, 211–215.

Shriver, M. D., Jin, L., Chakraborty, R. & Boerwinkle, E. (1993). VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134**, 983–993.

Slatkin, M. (1995). Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution* **12**, 473–480.

Stephan, W. & Cho, S. (1994). Possible role of natural selection in the formation of tandem-repetitive noncoding DNA. *Genetics* **136**, 333–341.

Stephan, W. & Langley, C. H. (1992). Evolutionary consequences of DNA mismatch inhibited repair opportunity. *Genetics* **132**, 567–574.

Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276.

Sutherlands, G. R. & Richards, R. I. (1995). Simple tandem DNA repeats and human genetic disease. *Proceedings of the National Academy of Sciences of the USA* **92**, 3636–3641.

Tachida, H. & Iizuka, M. (1992). Persistence of repeated sequences that evolve by replication slippage. *Genetics* **131**, 471–478.

Valdès, A. M., Slatkin, M. & Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**, 737–749.

Viard, F., Brémond, P., Labbo, R., Justy, F., Delay, B. & Jarne, P. (1996). Microsatellites and the genetics of highly selfing populations in the freshwater snail *Bulinus truncatus*. *Genetics* **142**, 1237–1247.

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

Weber, J. L. (1990). Informativness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics* **7**, 524–530.

Zhivotovski, L. A., Feldman, M. W. & Grishechkin, S. A. (1997). Biased mutations and microsatellite variation. *Molecular Biology and Evolution* **14**, 926–933.