Juxtaposed microsatellite systems as diagnostic markers for admixture: Theoretical aspects

Arnaud Estoup, Jean-Marie Cornuet, François Rousset, René Guyomard

MOLECULAR BIOLOGY AND EVOLUTION

Juxtaposed Microsatellite Systems as Diagnostic Markers for Admixture: Theoretical Aspects

Arnaud Estoup,†[1] *Jean-Marie Cornuet,*† *François Rousset,*‡ *and René Guyomard**

*Laboratoire de Génétique des Poissons, Institut National de Recherche Agronomique, Jouy-en-Josas, France; †Laboratoire de Modélisation et de Biologie Evolutive, Unité de Recherche de Lutte Biologique-Institut National de Recherche Agronomique, Montpellier, France; and ‡Laboratoire Génétique et Environnement, Centre National de Recherche Scientifique-Unité Mixte de Recherche, Montpellier, France

Two populations which have diverged from an ancestral population may come back into contact due to human action via stocking or introduction programs. We report here a method to measure genetic admixture in such situations based on juxtaposed microsatellite systems (JMSs). A JMS is composed of two microsatellite repeat arrays separated by a sequence of less than 200 bp. The advantage of a JMS stems from the superior genealogical information carried by the two microsatellite sites to that carried by just one. If five assumptions are fulfilled, JMSs provide reliable diagnostic markers which eliminate the need to know the genetic structure of the native population in the absence of admixture. Simulations show that optimal features at both microsatellite sites of the JMS are the occurrence of multistep mutations, moderately high mutation rates, and limited allele size constraints. Optimal demographic features include a relatively large number of generations since the separation of the alien and native populations and small population sizes, especially for the alien population. Substantial sampling of the alien population is also necessary.

Introduction

Natural or man-induced situations in which two genetically differentiated taxa hybridize and introgression occurs have received considerable interest in recent years and have been studied in all major groups of organisms (reviewed in Harrison 1993; Avise 1996). The situation considered in the present paper is that of two populations which have diverged from an ancestral population and have recently come back into contact due to human action via stocking or introduction programs. In this case, genetic admixture is asymmetric, as gene flow occurs only from the introduced population (hereinafter denoted the alien population) toward the native population. The study of such situations requires methods to accurately measure the proportion of admixture in native populations and in their individual genomes. Measurement of admixture may be achieved at both levels when diagnostic markers (markers with different alleles in the alien and native populations) are available (Harrison 1993; Avise 1996). For markers with overlapping allele distributions, one may compute various estimators of admixture coefficient at the population level based on the comparison of allele frequencies between the alien and the admixed populations, including genealogical information or not (e.g., Bertorelle and Excoffier 1998). However, all of the above methods require a good knowledge of the allele frequencies in the alien population and in the native population in the absence of admixture. This is possible for the alien population provided its origin is well documented and substantial sampling and marker analysis is achieved. In contrast, determining the allele frequencies prevailing in the native population in the absence of introduction is often not possible. Samples of the native population collected before introduction (temporal method) are often unavailable. Furthermore, even if preintroduction samples are available, temporal variation due to genetic drift and mutation may have substantially changed the genetic structure of the native population. This can be tentatively circumvented by determining allele frequencies in nearby populations, assuming that no alien individuals have been introduced and that allele frequencies are representative of those characterizing the admixed native population prior to admixture (geographical method) (e.g., Gyllensten et al. 1985; Giuffra, Guyomard, and Forneris 1996). However, because of the geographic structuring of genetic variation, these assumptions may lead to erroneous conclusions. With ''diagnostic markers,'' as defined by either the temporal or geographical methods, any allele documented in both the admixed native and the alien populations is considered to have an alien origin. However, these alleles may also have a native origin, with their similarities with alleles of the alien population resulting from identity by descent (ancestral alleles) or from homoplasy (alleles identical by state due to convergence, parallelism, or reversion events). Hence, usual allele assignment procedures using diagnostic markers potentially result in overestimation of admixture. The direction and extent of the bias associated with estimators of admixture coefficient based on the comparison of allele frequencies between the alien and the admixed populations are unpredictable.

Using microsatellite markers for admixture studies (Abernethy 1994; Gotelli et al. 1994; Roy et al. 1994; MacHugh et al. 1997; Goostrey et al. 1998) complicates the estimation of admixture rather than facilitating it. For a given divergence time, ancestral alleles are expected to be less frequent at highly mutating loci such as microsatellites than at low-mutating loci such as enzymes. However, a substantial amount of size homo-

[1] Present address: Department of Zoology, University of Queensland, Queensland, Australia.

Key words: diagnostic marker, genetic admixture, homoplasy, hybridization, linked markers, microsatellites.

Address for correspondence and reprints: Arnaud Estoup, Department of Zoology, University of Queensland, Queensland 4072, Australia. E-mail: aestoup@zoology.uq.edu.au.

Mol. Biol. Evol. 16(7):898–908. 1999

plasy is expected at these markers for at least three reasons: (1) length variation at microsatellites is usually due to stepwise variation in the repeat unit count (reviewed in Estoup and Cornuet 1999), (2) the high mutation rates of many microsatellite loci (Weber and Wong 1993) increase the likelihood of the same mutation occurring in parallel in different lineages, and (3) constraints act on allele size range, reducing the number of possible allelic states (Bowcock et al. 1994; Garza, Slatkin, and Freimer 1995; Nauta and Weissing 1996). The high levels of polymorphism of most microsatellites also result in potentially different allelic patterns among even closely related populations or successive cohorts of the same population (e.g., Estoup et al. 1998). This substantially complicates the assessment of genetic structure of the native population in the absence of admixture using the temporal or geographical methods.

Juxtaposed Microsatellite Systems as Diagnostic Markers for Admixture

A juxtaposed microsatellite system (JMS) is composed of two microsatellite repeat arrays separated by a sequence of less than 200 bp. None of the few studies on JMS polymorphism (Chakraborty et al. 1994; Pena et al. 1994; Dermitzakis et al. 1998) had the estimation of admixture coefficient as its objective. The advantage of a JMS for admixture study stems from the superior genealogical information carried by the two microsatellite sites to that carried by just one. More precisely, the information regarding sharing or not sharing an allele between the alien and the admixed native populations obtained at one microsatellite site is complemented by the sharing or nonsharing information at the second neighboring microsatellite site. This additional information is the cornerstone feature allowing a JMS to differentiate between alien alleles and naturally shared homoplasious or ancestral alleles and hence to be considered a reliable diagnostic marker of admixture. For both microsatellite sites of the JMS, we denote P (private alleles) any allele specific to the native population (i.e., not observed in the alien population) and S (shared alleles) any allele of the native population shared with the alien population. When all alleles of the alien population have been sampled, P alleles in the admixed native population are, by default, of native origin. However, S alleles may have an alien origin or a native origin, with their similarities with alleles of the alien population resulting either from identity by descent (ancestral alleles) or from homoplasy. For nonadmixed isolated native populations and a given divergence time, the probability that adjacent microsatellite sites both have homoplasious or ancestral alleles (SS haplotypes) is expected to be lower than that for a single microsatellite site (S allele). Under certain evolutionary assumptions that still need to be specified, this probability can be extremely low. Under these assumptions, PP, SP, and PS haplotypes sampled in an admixed population are of native origin, whereas SS haplotypes have a high probability of having an alien origin.

Table 1
Assumptions that Need to be Fulfilled for Valid Use of a Juxtaposed Microsatellite System (JMS) as Diagnostic Marker of Admixture

No.	Assumption
1...	The alien population does not contain any individuals/genes originating from the native population.
2...	No recombination has occurred within the JMS segments between two haplotypes, one (SS) of alien origin and one (PP) of native origin since admixture.
3...	No mutation at admixed alien haplotypes (SS) toward an allelic state specific to the native population (P) has occurred at either microsatellite site since admixture.
4...	Prior to admixture, the probability that a JMS haplotype in the native population is composed of shared alleles (S) at both microsatellite sites (SS haplotype) is extremely low.
5...	All alleles of the alien population have been sampled for both microsatellite sites composing the JMS.

NOTE.—P (private allele) stands for any single microsatellite site allele specific to the native population (i.e., not observed in the alien population), and S (shared allele) stands for any allele shared by the native and alien populations.

However, the proportion of SS haplotypes in the native population will be artificially increased if the alien population, even at low frequency, includes individuals originating from the native population. Moreover, two evolutionary events can potentially induce SS haplotypes of alien origin to be incorrectly diagnosed as native SP or PS haplotypes and cause an underestimation of admixture: (1) a recombination event occurring within the JMS segments between one SS haplotype of alien origin and one PP haplotype of native origin and (2) a mutation at one of the two microsatellite sites of admixed alien SS haplotype toward an allelic state specific to the native population (P). Thus, taking into account these considerations, we defined five assumptions (table 1) that need to be fulfilled for valid use of JMSs as diagnostic markers without any information on the genetic structure of the native population in the absence of admixture.

The interpretation of all possible categories of JMS genotypes assuming the five assumptions of table 1 is summarized in table 2. It is worth mentioning that the shared (S) or private (P) status of an allele in the native population is independent of its frequency in both the native and alien populations. SP, PS, and PP haplotypes are necessarily of native origin, and SS haplotypes have a high probability of corresponding to an introduced alien gene. Thus, an estimation of the admixture coefficient at a given JMS locus is simply the proportion of SS haplotypes in the admixed native population sample. Since microsatellite sites 1 and 2 are separated by less than 200 bp, the phase and, thus, the double-site haplotypes constituting the JMS genotype can be determined by PCR using the external primer of each microsatellite site and by comparing the sizes of the allelic fragments with the sizes of the allelic fragments obtained with two independent PCRs of each microsatellite site.

Validity of Assumptions 1–3

Assumption 1 (table 1) should be valid for numerous actual systems of admixture between alien and na-

Table 2
Interpretation of All Possible Genotypes at a Juxtaposed Microsatellite System (JMS) Under the Five Assumptions Presented in Table 1

Genotype at Microsatellite Site 1	Genotype at Microsatellite Site 2	JMS Genotype = Double-Site Haplotypes with Microsatellite Sites 1 and 2 in the First and Second Columns, Respectively	Number of JMS Haplotypes of Alien Origin	Complement of Information Assessing JMS Double-Site Phase
$^{P}/_{P}$	$^{P}/_{P}$	$^{PP}/_{PP}$: homozygous native JMS	0	No
$^{S}/_{S}$	$^{S}/_{S}$	$^{SS}/_{SS}$: homozygous alien JMS	2	No
$^{P}/_{S}$	$^{P}/_{S}$	$^{PP}/_{SS}$: native-alien hybrid JMS or	1	Yes: differentiation between the two possible genotypes
		$^{PS}/_{SP}$: homozygous native JMS with one homoplasious/ancestral allele at both sites	or 0	
$^{P}/_{P}$	$^{P}/_{S}$	$^{PP}/_{PS}$: homozygous native JMS with one homoplasious/ancestral allele at site 2*	0	No
$^{P}/_{P}$	$^{S}/_{S}$	$^{PS}/_{PS}$: homozygous native JMS with two homoplasious/ancestral alleles at site 2*	0	No
$^{P}/_{S}$	$^{S}/_{S}$	$^{PS}/_{SS}$: native–alien hybrid JMS with one homoplasious/ancestral allele at site 2*	1	Yes: determination of the homoplasious/ancestral allele at site 2

NOTE.—For both microsatellite sites composing the JMS, P (private allele) stands for any allele specific to the native population (i.e., not observed in the alien population), and S (shared allele) stands for any allele shared by the native and alien populations. * = reverse sites 1 and 2 for reciprocal cases.

tive taxa (see *Discussion*). The second assumption (table 1) should be valid when analyzing microsatellites separated by a short sequence, as for a JMS. Assuming that 1 cM corresponds to ca. 1×10^6 bp as in humans (Mirsky and Riss 1951; Dib et al. 1996), one recombination between two sites 100 bp apart is expected every 1×10^6 generations. Hence, recombination should not represent a major problem for most alien–native case studies, except if microsatellites represent recombinational hot spots. Although the occurrence of rare interallelic events involving crossover or gene conversion cannot be ruled out, indirect evidence suggests that most repeat number variation at microsatellites corresponds to intraallelic events involving a replication slippage mechanism (Levinson and Gutman 1987; Schlötterer and Tautz 1992). Finally, although more difficult to quantify, the third assumption (table 1) should be met if the mutation rates at the microsatellite sites are not too high and if the introduction of the alien population is relatively recent, such that mutations occurring between introduction and sampling times can be neglected along the ancestral lines of the sampled genes.

Validity of Assumption 4

At mutation-drift equilibrium, the validity of assumption 4 (table 1) depends on two population parameters, the population size and the time since the separation of the alien and the native populations before admixture. It also depends on two marker parameters, the mutation rate and the mutation model. We used a simulation approach to assess the influence of these parameters on the proportion of SS haplotypes in a nonadmixed native population. Simulations of the coalescent process were performed according to Simonsen, Churchill, and Aquadro (1995), assuming no gene exchange between the two populations since they separated and no population size fluctuation over time. The demographic and marker conditions assumed in standard simulation conditions are equal population sizes of 1,000 individuals and mutation rates of 5×10^{-4} for both populations and microsatellite sites, the latter evolving under a stepwise mutation model (SMM; Kimura and Ohta 1978) with no constraint on allele size. The number of JMS haplotypes sampled after each iteration was 60 for the native population (30 diploid individuals) and 900 for the alien population (450 diploid individuals). The large size of the alien sample is justified by assumption 5 (table 1; for details, see *Validity of Assumption 5*). Proportions of SS, PP, SP, and PS haplotypes in the native sample were computed at each replicate. An estimate of the expectation of these proportions was computed from 1,000 iterations (E[SS], E[PP], E[SP], and E[PS]). An additional statistic, denoted Abs(SS), was computed. It is defined as the proportion of replicates for which SS haplotypes were absent from the 60 genes sampled in the native population. Abs(SS) can be interpreted as: (1) (considering a single native population) the proportion of JMSs with no SS haplotypes among the 60 genes sampled in this population and, thus, the proportion of strictly diagnostic JMS markers at this sampling level, or (2) (considering a single JMS) the proportion of native populations (replicates of the simulated evolutionary process) for which this JMS will behave as a strictly diagnostic marker. Because one of

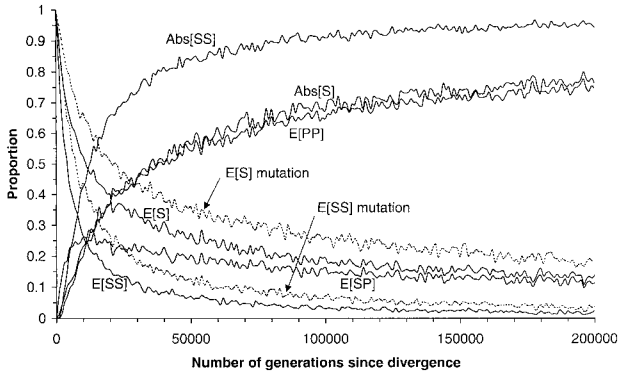

FIG. 1.—Relationship between the divergence time and *E*[SS, SP, and PP] or Abs[SS, SP, and PP] statistics for a JMS, as well as for a single-site microsatellite locus (*E*[S] and Abs[S]). Population and marker parameters correspond to standard simulation conditions, i.e., $N = 1{,}000$ for both populations and $\mu = 5 \times 10^{-4}$ for both microsatellite sites, with the latter following an SMM with no allele size constraint. The dashed curves represent the contribution of mutations to the decline of *E*[SS] and *E*[S] with time (see text for details).

the main objectives of this study was to evaluate the informational gain associated with the use of a JMS as compared with a usual single-site microsatellite locus, similar statistics were computed independently for each microsatellite site composing the JMS (*E*[S1] and Abs[S1] for site 1 and *E*[S2] and Abs[S2] for site 2). Note that S1 and SP statistics are, respectively, equivalent to S2 and PS statistics when both microsatellite sites of the JMS have the same mutation rate and model. In this case, only S1 and SP values are presented in figures and tables.

Population Parameters

Divergence Time

The variation of SS and S statistics for divergence times ranging from 0 to 200,000 generations (step of 1,000 generations) are given in figure 1 for standard simulation conditions. As expected, *E*[SS] (Abs[SS]) values decrease (increase) with the divergence time much more rapidly than do *E*[S] (Abs[S]) values. Since reliable diagnostic markers call for low (high) values of *E*[SS] or *E*[S] (Abs[SS] or Abs[S]), the use of the JMS potentially represents a substantial gain regarding the range of divergence time that can be analyzed as compared with that using a single-site microsatellite locus. For instance, under standard conditions, a divergence time of 150,000 generations is associated with high Abs[SS] (95%) and low *E*[SS] (2.5%) values, while comparatively low Abs[S] (74%) and high *E*[S] (16%) values make the use of a single-site microsatellite locus less appropriate. This result holds for any demographic and marker conditions studied in this paper.

Population Size

E[SS] is the average fraction of SS haplotypes in the native population, which is also the probability that a single haplotype in the native population is SS. This probability depends on the mutation events occurring in the ancestral lineage of the native haplotype. These mutation events are independent of the size of the native

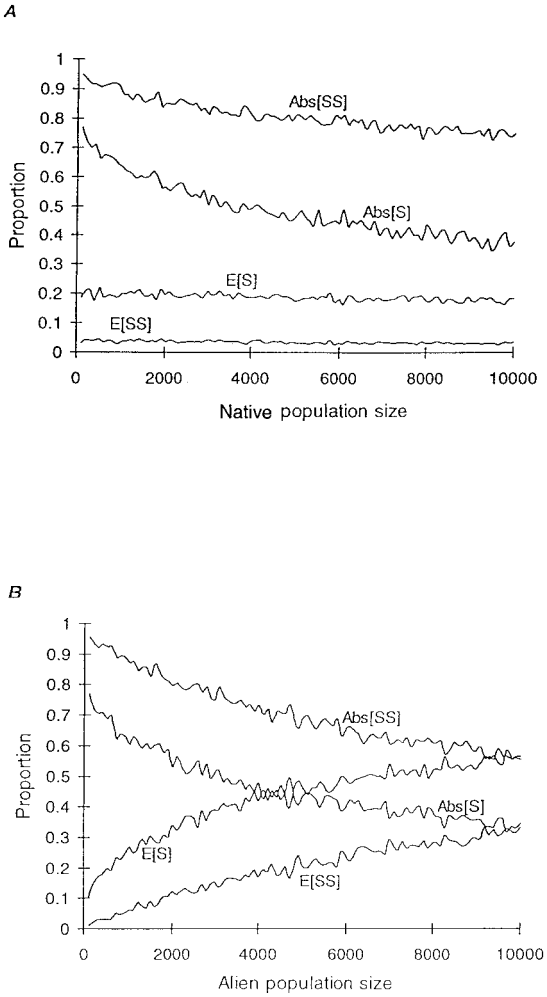

FIG. 2.—Relationship between the population size of the alien (N_a) and native (N_n) populations and *E*[SS and S] or Abs[SS and S] statistics. N_n or N_a values vary by steps of 100 from 100 to 10,000 with values of N_a (*A*) and N_n (*B*) kept constant at 500. The divergence time was kept at 50,000 generations. Other marker parameters are as in the standard simulation conditions.

population (N_n). However, *E*[SS] is expected to increase with the size of the alien population (N_a), since larger alien populations will comprise more different double-site haplotypes. Regarding the statistic Abs[SS], it is worth noting that $\text{Abs[SS]} = E[(1 - f_{SS})^n]$, where n is the number of genes sampled in the native population and f_{SS} is the frequency of SS haplotypes in the native population. f_{SS} is a random variable whose distribution depends on the size of both the alien and the native populations (results not shown). Thus, in contrast to *E*[SS], Abs[SS] is expected to be also dependent on N_n. These expectations were confirmed and detailed through simulation by varying N_n (N_a) values by steps of 100 from 100 to 10,000 diploid individuals and by keeping N_a (N_n) constant at 500. Figure 2*A* shows that for a constant N_a value, whereas *E*[SS] and *E*[S] values remained similar for different N_n values, Abs[SS] and Abs[S] values decreased when N_n values increased. For a constant N_n value, *E*[SS] and *E*[S] (Abs[SS] and

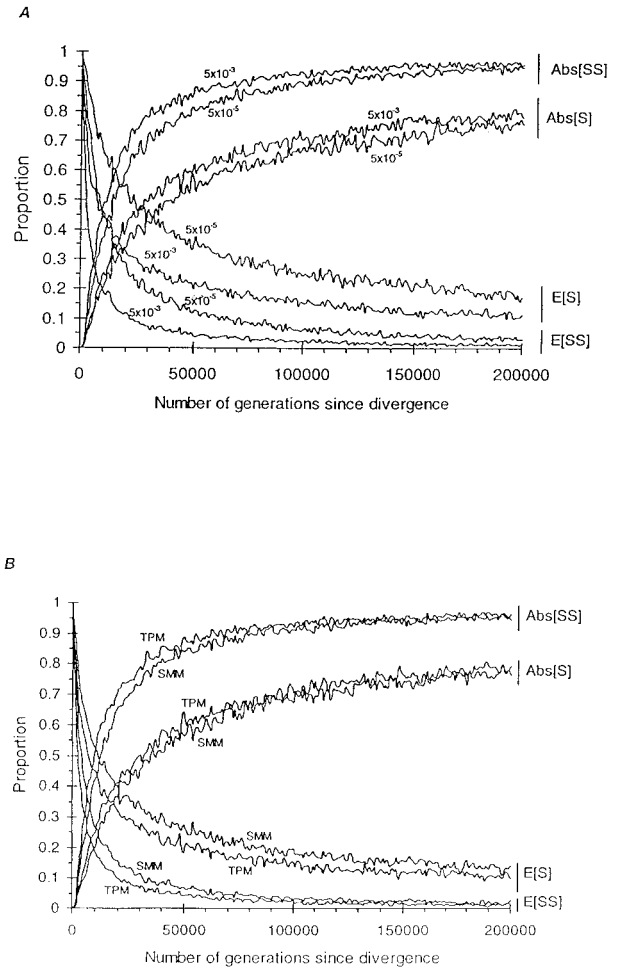

FIG. 3.—Effect of the mutation rate ($\mu = 5 \times 10^{-5}$ and 5×10^{-3}) on *E*[SS and S] and Abs[SS and S] statistics (*A*) and effect of large mutation steps (*B*). TPM parameters are $P = 0.674$ and $\alpha = 0.5$. Other population and marker parameters are as in the standard simulation conditions.

Abs[S]) values increased (decreased) when N_a values increased (fig. 2*B*). Hence, JMSs should provide reliable diagnostic markers even for moderately short divergence times when populations (especially alien populations) of small sizes are studied. For instance, Abs[SS] and E[SS] values will be $\geq$ 95% and $\leq$ 2.5%, respectively, for divergence times $\geq$ 25,000 generations if N_a and N_n both equal 100 individuals.

Marker Parameters

Mutation Rate

The influence of the mutation rate was studied by computing SS and S statistics for high (5×10^{-3}) and low (5×10^{-5}) mutation rates, according to the range of microsatellite mutation rate (10^{-2}–10^{-5}; Weber and Wong 1993). For any divergence time, higher mutation rate corresponded to lower values of *E*[SS] and *E*[S] and higher values of Abs[SS] and Abs[S] (fig. 3*A*). Thus, JMSs composed of microsatellites with high mutation rates should be more appropriate for providing reliable diagnostic markers than JMSs composed of microsatellites with low mutation rates (but see the sections dealing with the validity of assumption 3 and allele size constraints).

Mutation Model

Effect of large mutation steps: Pedigree analysis, population studies, and allele sequencing have shown that, although rarer, mutation steps larger than a single repeat unit occurred at least at some microsatellites (reviewed in Estoup and Cornuet 1999). This evolutionary feature can be modeled by the two-phase model (TPM; DiRienzo et al. 1994). Maximum-likelihood estimates (MLEs) of the TPM parameters P and α can be computed from a sample of mutations for which the size change in repeat number is known (see appendix). To our knowledge, sufficiently large sets of mutations to allow precise parameterization of the TPM over a large number of microsatellite loci are still unpublished. The only published large mutation sample is composed of 43 germ line mutations detected at a single tetranucleotide locus (Primmer et al. 1998). Using Primmer et al.'s (1998) mutation size distribution, we found an MLE of $P = 0.674$ and $\alpha = 0.5$. Figure 3*B* compares SS and S statistic values assuming a TPM with $P = 0.674$ and $\alpha = 0.5$ with those obtained under an SMM. TPM values of *E*[SS] and *E*[S] are always lower and TPM values of Abs[SS] and Abs[S] are always higher than those obtained under an SMM. Hence, the occurrence of mutation steps larger than a single repeat unit favors the use of JMSs as diagnostic markers in the sense that, all other things being equal, low *E*[SS] values and high Abs[SS] values are obtained for lower divergence times (but see section on allele size constraints). This trend increases when the proportion and/or the mean size of large mutation step increases (results not shown).

Allele size constraints: This evolutionary feature was incorporated into our simulation by imposing reflecting boundaries on the allele size range (Feldman et al. 1997; Pollock et al. 1998). We varied the range of contiguous allelic states (R) from 10 to 100 (see *Discussion*), with identical ranges for both microsatellites composing the JMS. No significant effect on Abs[SS] or *E*[SS] values was observed for R values $\geq$ 50, with other simulation parameters as in standard simulation conditions. For a smaller value of R, e.g., $R = 20$, *E*[SS] and *E*[S] values were higher and Abs[SS] and Abs[S] values were lower than those obtained with $R = 50$ (fig. 4*A*). Under our simulation conditions, the difference between $R = 20$ and other curves appear after ca. 70,000 generations of divergence.

Low allelic ranges (e.g., $R < 20$ and other simulation parameters fixed as in standard conditions) converged to moderately low asymptotic Abs[SS] values ($<$0.90) and high asymptotic *E*[SS] values ($>$0.05), making the use of JMSs inappropriate for admixture studies between even highly divergent populations (table 3). This trend is stronger for a single-site microsatellite locus, since asymptotic Abs[S] and *E*[S] values for $R <$ 20 were substantially lower ($<$0.64) and higher ($>$0.22), respectively. Interestingly, R values between 30 and 50 (other simulation parameters as in standard conditions)

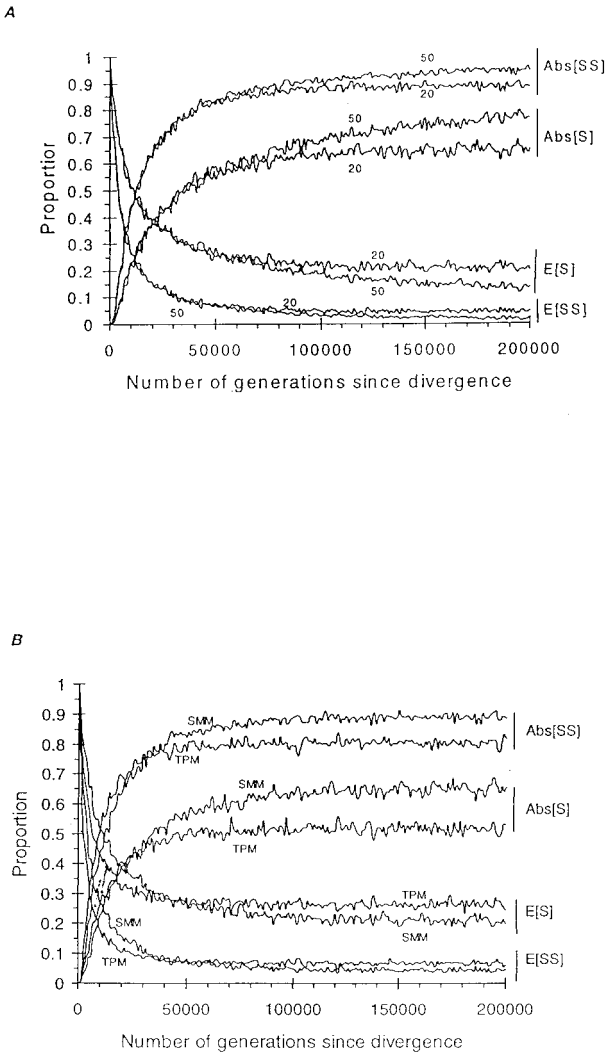

FIG. 4.—Effect of allele size constraints ($R = 20$ and 50) on E[SS and S] and Abs[SS and S] statistics (fig. 4*A*). Effect of the mutation model assuming allele size constraint with $R = 20$ (fig. 4*B*). TPM parameters are $P = 0.674$ and $\alpha = 0.5$. Other population and marker parameters are as in the standard simulation conditions.

lead to high asymptotic Abs[SS] values (0.95–0.98) and low asymptotic E[SS] values (0.01–0.02), and to substantially lower asymptotic Abs[S] values (0.75–0.85) and higher asymptotic E[S] values (0.10–0.15) (table 3). Hence, with these simulation conditions, single-site microsatellite loci are much less likely to provide reliable diagnostic markers than are JMSs, even for highly divergent populations. Simulation results also showed that, all other things being equal, a TPM results in a more rapid convergence toward lower (higher) asymptotic values of E[SS] (Abs[SS]) than does an SMM (fig. 4*B*). Note that the difference between the asymptote values under a TPM and those under an SMM decreases when R increases (table 3). Moreover, Abs[SS] and Abs[S] (E[SS] and E[S]) asymptotic values decrease (increase) when the mutation rate and/or the population size increases (table 4).

Table 3
Asymptotic Values for Divergent Times $\rightarrow \infty$ of E[SS], E[S], Abs[SS], and Abs[S] Under Allele Size Constraints Corresponding to Different Ranges of Allele Size (R) and Assuming an SMM or a TPM with $P = 0.674$ and $\alpha = 0.5$

	SMM				TPM			
R	E[SS]	E[S]	Abs[SS]	Abs[S]	E[SS]	E[S]	Abs[SS]	Abs[S]
10	0.143	0.381	0.684	0.404	0.209	0.455	0.493	0.250
20	0.049	0.221	0.897	0.637	0.067	0.260	0.807	0.513
30	0.022	0.148	0.950	0.753	0.036	0.193	0.896	0.649
40	0.013	0.116	0.971	0.808	0.018	0.138	0.945	0.725
50	0.009	0.095	0.979	0.846	0.013	0.117	0.956	0.774
75	0.004	0.064	0.991	0.892	0.005	0.081	0.982	0.846
100	0.002	0.046	0.993	0.922	0.002	0.056	0.991	0.890
∞	0.000	0.000	1.000	1.000	0.000	0.000	1.000	1.000

NOTE.—Other simulation parameters are $N_n = N_a = 1{,}000$ and $\mu = 5 \times 10^{-4}$ for both microsatellite sites composing the JMS.

General Considerations on the Expectation of SS and S Haplotypes

Two factors interact to control the evolution of E[SS] and E[S] following the separation of populations. First, mutations change haplotypes within both the alien and native populations. Second, coalescences occurring within the alien population reduce the diversity of ancestral haplotypes in the alien population. Note that, since assessing the status of single-site and JMS haplotypes consists of comparing each haplotype of the native population to all haplotypes of the alien population, coalescent events within the native population will have no effect on E[SS] and E[S]. In order to independently assess the contributions of mutation and coalescence to the decline of E[SS] and E[S] with time, E[SS] and E[S] values were computed between the present native population and the ancestral alien population (the alien population just before separation) for different divergence times. Since these simulations do not take into account the mutations occurring in the alien populations, they give a lower bound of the effect of mutation on E[SS] and E[S]. However, the effect of alien mutation is expected to be limited, since all copies of a given haplotype have to be removed by mutation to have an effect on E[SS] or E[S]. The effects of mutation on E[SS] and E[S] are presented in figure 1 for standard simulation conditions. The contribution of coalescent events to the decline of E[SS] and E[S] can be estimated

Table 4
Effect of the Mutation Rate (μ) and Population Size (N) on Asymptotic Values for Divergent Times $\rightarrow \infty$ of E[SS], E[S], Abs[SS], and Abs[S] Under an SMM with Allele Size Constraints Corresponding to $R = 30$

μ	N	E[SS]	E[S]	Abs[SS]	Abs[S]
5×10^{-3} . .	1,000	0.081	0.287	0.839	0.584
5×10^{-4} . .	1,000	0.022	0.148	0.950	0.753
5×10^{-5} . .	1,000	0.004	0.067	0.993	0.910
5×10^{-4} . .	100	0.005	0.066	0.992	0.911
	1,000	0.022	0.148	0.950	0.753
	10,000	0.126	0.353	0.669	0.389

by looking at the differences between the curve obtained as explained above and the usual curve, with the latter including the effects of both coalescence and mutation. Figure 1 shows that coalescence events substantially contribute to the rapid decline of *E*[SS] and *E*[S] following population separation. The slower decline that then ensues essentially results from the accumulation of mutations pushing apart the distribution of allele sizes in the two populations. Note that the relative contribution of mutation is larger for *E*[SS] than for *E*[S]. This is expected, as mutations occurring at either of the two microsatellite sites change JMS haplotypes within populations. Similar patterns were observed for other marker and population parameters (not shown).

Since the mutation processes at each microsatellite site of a JMS are assumed to be independent, *E*[SS], *E*[PP], *E*[SP], and *E*[PS] are expected to be equal to the products between *E*[S1] and *E*[S2], *E*[P1] and *E*[P2], *E*[S1] and *E*[P2], and *E*[P1] and *E*[S2], respectively, whatever the initial composition of the ancestral population. Simulation results confirmed these expectations for all simulation conditions studied in this paper (not shown).

Validity of Assumption 5

Assumption 5 (table 1) mainly requires a sufficiently large sampling of the alien population. For different sizes of the alien gene sample, we computed *E*[SS] as well as the proportion of the replicates for which only SS haplotypes were present in the 60 genes sampled in the native population for a divergence time of zero generations (Tot[SS]). *E*[SS] and Tot[SS] are equal to 1 if all alien alleles have been sampled. Figure 5*A* and *B* show that an insufficiently large sampling of the alien population results in an underestimation of *E*[SS] and Tot[SS] values. The larger the $M = 4N\mu$ values, the larger is this bias. Assuming an SMM with no allele size constraint, all tested *M* values resulted in *E*[SS] values of >0.99 when the number of genes sampled in the alien population (n_a) was ≥ 180. However, because of large variance in SS proportions, Tot[SS] values responded much more drastically to variation in the number of sampled alien genes. Tot[SS] values were indeed ≥ 0.95 if $n_a \geq 280$, 480, and 880 for $M = 0.2$, 2 (standard conditions), and 20, respectively. Lowered *E*[SS] and Tot[SS] values were obtained under a TPM (fig. 5*A* and *B*). In contrast, allele size constraints with $R \geq 20$ did not change *E*[SS] and Tot[SS] values (results not shown).

Nine hundred gene copies were sampled in our simulations. For all evolutionary conditions studied, a simulated divergence time of zero generations gave *E*[SS] values of >0.999 and Tot[SS] values of $\geq 95\%$ (fig. 5*A* and *B* and results not shown). Note also that if one assumes a binomial distribution of gene count for any allelic type, any allele with a frequency of ≥ 0.0034 has a probability of $<5\%$ of not being sampled in a sample of 900 gene copies. These simulation results and arguments both indicate that the SS, PP, SP, and PS statuses of JMS haplotypes were precisely assessed in our simulations.

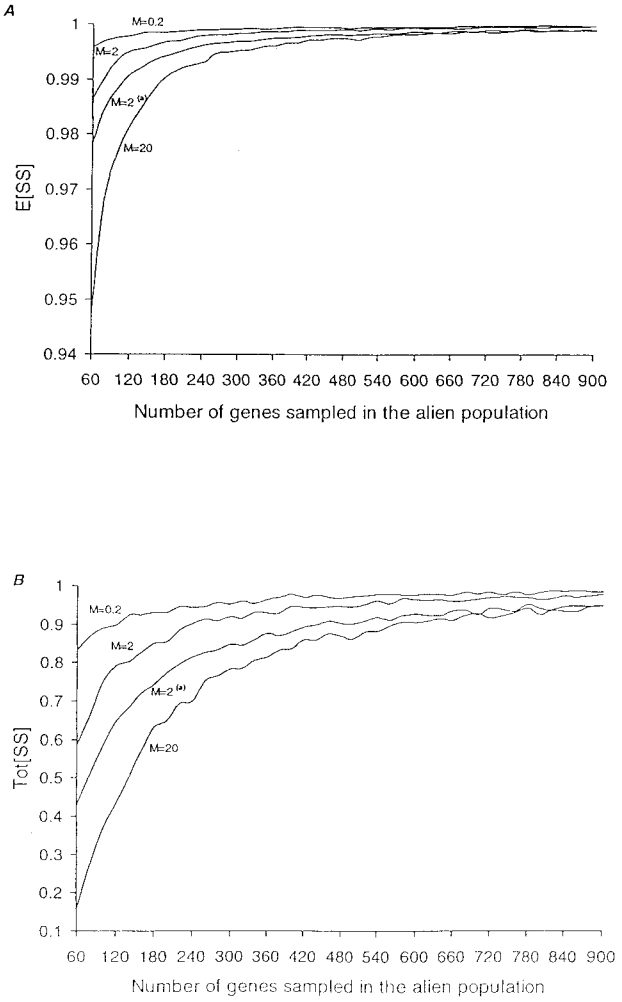

FIG. 5.—Effect of the number of genes sampled in the alien population on (*A*) *E*[SS] and (*B*) Tot[SS] statistics for different $M = 4N\mu$ values and assuming an SMM or a TPM(a) with $P = 0.674$ and $\alpha = 0.5$. The divergence time was kept constant at zero generations.

Estimation of Admixture Coefficient

When assumptions 1–3 are fulfilled and when relevant marker and demographic parameters are available, simulation curves of the present study can provide guidelines to assess the ability of JMSs to fulfill assumption 4 (table 1) for a given admixture situation. Allele size constraints, mutation rate, population size, and separation time can be roughly empirically estimated from nonadmixed populations (reviewed in Estoup and Angers [1998]). Besides these parameter estimates, we advise preliminarily testing a set of JMSs in the context of the admixture situation that one aims at studying. This could be achieved by genotyping an alien population sample of moderate size and a few nonadmixed native population samples for these JMSs. SS haplotypes should be absent from these preliminary tests. A larger sampling and genotyping of the alien population is then necessary to fulfill assumption 5 (table 1). Without any information on the relevant marker and demographic parameters, and if no nonadmixed populations are available, one may simply observe some level of

haplotype sharing which could be due to a low separation time or a larger separation time with admixture. In this case, it becomes difficult to assess the validity of assumption 4 (table 1), such that using JMSs as strictly diagnostic markers is uncertain. For any demographic and marker parameters, the only reliable information on allele origin comes from PS, SP, and PP haplotypes, which are necessarily of native origin. Thus, the sums of the proportions of PS, SP, and PP haplotypes may be computed as a potential underestimate of the actual nonadmixed gene fraction. According to our simulations, such estimates will, in any case, be more precise than any estimate based on a single microsatellite site.

If the five assumptions of table 1 (especially assumption 4) are fulfilled, the status of SP, PS, PP, and SS haplotypes in an admixed population can be determined at one or several diagnostic JMSs, and an estimation of the admixture coefficient is given at each JMS by the proportion of SS haplotypes in the sample. Confidence intervals on each JMS admixture coefficient can be obtained assuming a binomial distribution of SS haplotype count or by bootstrapping over haplotypes or individuals. A mean admixture coefficient ($\overline{AC}$) can be computed as $\overline{AC} = (\Sigma_{j=1}^{k} \mathrm{SS}_j)/(\Sigma_{j=1}^{k} \mathrm{SS}_j + \mathrm{PP}_j + \mathrm{SP}_j + \mathrm{PS}_j)$ for k genetically independent JMSs, and confidence intervals on the mean can be obtained by bootstrapping over JMSs. For low numbers of JMSs, confidence intervals on $\overline{AC}$ can be obtained by assuming a multinomial distribution of SS haplotype count with the same probability for each locus (Sokal and Rohlf 1995). If individual genomes are considered, the mean admixture coefficient for a given genome can be computed as $AC = (1/2k)\Sigma_{j=1}^{k} \mathrm{SS}_j$ for k genetically independent JMSs, and confidence intervals can be obtained as above.

Comparison between AC values of two (or more) JMSs can be achieved using exact tests on two-way contingency tables (Sprent 1989). Comparison between $\overline{AC}$ values computed at either the population or the individual level in different admixed populations or individuals can be achieved by summing the two classes of haplotypes (SS and non-SS haplotypes) over loci in each population or individual and using exact tests on two-way contingency tables (Sprent 1989) or, alternatively, using a model I ANOVA or a t-test (Sokal and Rohlf 1995).

Discussion

Isolation of JMSs

Although both in situ hybridization and genetic mapping have revealed a relatively even distribution of microsatellites over chromosomes, the frequent association of several microsatellite sequences in the same cloned insert is indicative of a clustering of microsatellites in bees (Estoup et al. 1993; Thoren, Paxton, and Estoup 1995). A similar trend was suggested for mammals by high density maps of the human and mouse genomes (Dietrich et al. 1994; Dib et al. 1996). Hence, the isolation of JMSs using classical cloning methods should be possible, if not easy, provided that a sufficiently high density of microsatellites exits in the genome under study. It should be possible to increase the probability of isolating and sequencing JMSs by screening clones with a mixture of probes representing different repeat motifs and by preferentially sequencing the positive clones characterized by a strong hybridization signal and a long insert. For instance, JMSs were found in ca. 15% of microsatellite clones with intense positive signals and relatively large inserts (450–1,000 bp) in the genome of salmonid fishes. For any genome, and especially for those with low microsatellite density, JMSs could be specifically isolated without the need for classical genomic library construction by using a random amplified microsatellite system (RAMS; Browne and Litt 1992; Hantula, Dusabenyagasani, and Hamelin 1996). This method is based on the PCR and uses primers containing microsatellite sequences and degenerate anchors at the 5′ end. It directly and essentially screens for juxtaposed microsatellites, although compound microsatellites may also be isolated.

Marker and Demographic Parameters

Simulations have shown that constraint on allele sizes is one of the most influential factors on E[SS] and Abs[SS] values and, hence, on the ability of a JMS to provide reliable diagnostic markers. Allele size constraints were included in our simulations by imposing reflecting boundaries on the allele size range. This mechanism presents several advantages, namely great simplicity of programming, previous theoretical developments (e.g., Goldstein et al. 1995; Feldman et al. 1997), and the possibility of estimating boundary values from empirical data (Pollock et al. 1998). Because a mechanism based on reflecting boundaries has little empirical support, several other mechanisms have recently received attention (reviewed in Amos [1999] and in Estoup and Cornuet [1999]). Although anecdotal support has been documented for some of them, further work is needed to select and parameterize one of these alternative mechanisms. The range of contiguous allelic states (R) considered in our simulations (10–100) is consistent with empirical data on repeat numbers at microsatellites in various species (Garza, Slatkin, and Freimer 1995; Goldstein and Pollock 1997) and with R values estimated using the method of Pollock et al. (1998) applied to empirical microsatellite data for the brown trout and other salmonids.

The mutation models considered in our simulations do not include two factors which have been identified as being relevant to the evolution of microsatellites: (1) a complex dependence of the mutation process on repeat count and purity of alleles, and (2) that mutations at microsatellites involve more gains than losses of repeats (reviewed in Estoup and Cornuet 1999). However, we currently do not have the data to accurately parameterize and adequately include these evolutionary factors in theoretical mutation models. It is unlikely that adding these factors to our simulations would significantly change the major conclusions of this study.

The chosen window of divergence times (0–200,000 generations) corresponds to populations belonging to the same species or to closely related species. Much larger divergence times were not, or, at best, were

punctually studied because they would correspond to evolutionary situations for which the gain expected from a JMS would be usually low as compared to a single-site microsatellite locus. However, a substantial gain of JMSs still exists for highly divergent populations when low-mutating markers and/or large population sizes are simulated, as well as for particular ranges of allele size constraints resulting in low (high) E[SS] (Abs[SS]) and rather high (low) E[S] (Abs[S]) asymptotic values.

All simulations of this work assumed that population sizes are constant over time and that no gene exchange occurred between the two populations after they separated and before admixture. The exact influence of population size fluctuation(s) and past migration on E[SS] and Abs[SS] still needs to be studied. However, according to the simulation results on the effect of population size in the present study, qualitative predictions are likely to be, all other things being equal, a decrease or an increase of E[SS] if a bottleneck or an expansion occurred in the alien population, respectively. However, size fluctuation of the native population should not affect E[SS]. In contrast, Abs[SS] should be affected by a bottleneck and an expansion, whatever the population in which it occurred. Finally, uni- and bidirectional past migrations are both expected to increase (decrease) E[SS] (Abs[SS]).

Relationships Between JMS and Single-Site Haplotypes

In our simulations, we assumed that the mutation process at one site of the JMS was independent of the mutation process at the other site. Pena et al. (1994) observed positive allele size association at a JMS when considering double-site haplotype occurrences in a human population. In contrast, negative covariance in repeat number was found between the two sites of a JMS genotyped in a fish population (Dermitzakis et al. 1998). Both studies suggest that JMS may not evolve independently, but the evolutionary processes potentially driving the pattern of intersite disequilibrium are still unclear and appear to change substantially among JMSs. Further studies should be conducted on additional JMSs and on different populations and species before considering these results general evolutionary features of JMSs.

Alleles of type S correspond to ancestral alleles or to homoplasious alleles. The fraction of ancestral alleles is expected to decrease with divergence time as a result of genetic drift and mutation. Analytical and simulation results for two diverging populations have shown that for a wide range of evolutionary conditions compatible with those of microsatellites, the fraction of shared alleles identical by descent between two populations rapidly decreased so that most S alleles corresponded to homoplasious alleles (unpublished data). For example, if $M = 4N\mu = 1$ (e.g., $\mu = 5 \times 10^{-4}$ and $N = 500$) for both the alien and the native populations, 95% of S alleles are homoplasious after less than $4N$ generations under an SMM and after less than $3N$ generations under a TPM with $P = 0.674$ and $\alpha = 0.5$ (unpublished data). The number of generations required to fulfill assumption 4 (table 1) is substantially larger than the above numbers. Therefore, a diagnostic JMS will essentially differentiate between introduced and homoplasious alleles.

Application of JMSs

The classical cases of application of JMSs as diagnostic markers involve native population(s) which were recently admixed with genetically differentiated alien individuals introduced by man in the context of stocking or introduction programs. Numerous potential case studies exist, concerning a wide variety of organisms. As examples, we can cite, for fish, the stocking of native brown trout (*S. trutta*) populations with hatchery trout from the same or different subspecies (e.g., Giuffra, Guyomard, and Forneris [1996]); for insects, the issue of the Africanized honeybee *Apis mellifera* in South and Central America or the introduction of foreign honeybee queens in west European apiaries by beekeepers (Lobo, Del Lama, and Mestriner 1989; Rinderer et al. 1991; Estoup et al. 1995); and for mammals, the admixture in Scotland of native red deer (*Cervus elaphus*) populations with japonese sika deer (*Cervus nippon*) (Abernethy 1994) or the recent admixture of Ethiopian wolf populations (*Canis simensis*) with domestic dogs (Gotelli et al. 1994).

The use of JMSs could tentatively be extended to the study of recent hybrid zones induced by habitat changes caused by human development. For instance, a potentially appropriate situation could be the hybridization of three closely related North American canid species, the gray wolf (*Canis lupus*), the coyote (*Canis latrans*), and the red wolf (*Canis rufus*), among which substantial overlap of microsatellite allele size distribution has been observed (Roy et al. 1994). Note that for such natural hybrid zones, the general procedure of using JMSs is more complex than it is for the previous unidirectional admixture case studies. Since either taxon can be considered the alien population, the study of a hybrid zone would first require a large sampling of the gene diversity of both taxa outside the hybrid zone. Subsequently, individuals from the hybrid zone would be analyzed by considering alternatively each taxon as the alien population.

Finally, it has been argued that protein-coding loci may introgress more slowly than neutral nuclear markers such as anonymous restriction fragment length polymorphisms or microsatellite loci because of selection constraints acting on functional markers in oysters (Karl and Avise 1992; but see Hare and Avise 1996) or in brown trout (Poteaux, Bonhomme, and Berrebi 1998). Testing this hypothesis requires that homoplasious/ancestral alleles be clearly identified and withdrawn from the computation. JMSs could be appropriate markers to fulfill this requirement, at least for particular evolutionary situations.

Acknowledgments

We thank Marie-Charlotte Anstett, Yannis Michalakis, Craig Primmer, Julie Turgeon, Ross Crozier, and two anonymous reviewers for constructive comments.

This work was partly supported by the program ''Recherches méthodologiques pour l'amélioration des processus de gestion et de conservation des ressources génétiques animales, végétales et microbiennes'' of the Bureau des Ressources Génétiques for A.E. and R.G.

APPENDIX

In the TPM (DiRienzo et al. 1994), mutations introduce a gain/loss of X repeats. With probability P, X is equal to one and with probability $1 - P$, X follows a geometric distribution, with parameter α defined as $\text{Proba}(X = i) = (1 - \alpha)\alpha^{i-1}$, with variance $V(X) = \alpha/(1 - \alpha)^2$ and expectation $E(X) = 1/(1 - \alpha)$. We have derived MLE expressions of P and α that can be computed from a sample of mutations for which the size change in repeat number is known.

The likelihood of a sample $\mathbf{n} = (n_1, n_2, \ldots, n_k, \ldots)$, with n_i being the number of mutation events corresponding to a gain or a loss of i repeats, is, under a TPM,

$$L(\mathbf{n}; P; \alpha) = \frac{n!}{n_1! \ldots n_k! \ldots}[1 - (1 - P)\alpha]^{n_1} \times \prod_{i \geq 2} [(1 - P)(1 - \alpha)\alpha^{i-1}]^{n_i}.$$

Solving $(\partial \log L)/\partial P = 0$ and $(\partial \log L)/\partial \alpha = 0$ gives the MLE

$$\hat{P} = 1 - \frac{\sum_{i>1} n_i \sum_{i>1} (i - 1)n_i}{n \sum_{i>2} (i - 2)n_i} \quad \text{and} \quad \hat{\alpha} = \frac{\sum_{i>2} (i - 2)n_i}{\sum_{i>1} (i - 1)n_i},$$

except in some special cases in which one of the partial derivatives cannot be zero for possible parameter values. These cases are:

1. Only $n_1 > 0 \rightarrow \hat{P} = 1$, no estimate for α.
2. Only $n_2 > 0 \rightarrow \hat{P} = 0$, $\hat{\alpha} = 1/2$.
3. Only n_1 and $n_2 > 0 \rightarrow \hat{P} = 0$, $\hat{\alpha} = n_2/(n_1 + n_2)$.

LITERATURE CITED

ABERNETHY, K. 1994. The establishment of a hybrid zone between red and sika deer (genus *Cervus*). Mol. Ecol. **3**:551–562.

AMOS, W. 1999. A comparative approach to the study of microsatellite evolution. *In* D. B. GOLDSTEIN and C. SCHLÖTTERER, eds. Microsatellites: evolution and applications. Oxford University Press, Oxford (in press).

AVISE, J. C. 1996. Introduction: the scope of conservation genetics. Pp. 1–9 *in* J. C. AVISE and J. L. HAMRICK, eds. Conservation genetics: case histories from Nature. Chapman and Hall, New York.

BERTORELLE, G., and L. EXCOFFIER. 1998. Inferring admixture proportions from molecular data. Mol. Biol. Evol. **15**:1298–1311.

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J. R. KIDD, and L. L. CAVALLI-SFORZA. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368**:455–457.

BROWNE, D. L., and M. LITT. 1992. Characterization of $(CA)_n$ microsatellites with degenerate sequencing primers. Nucleic Acids Res. **20**:141.

CHAKRABORTY, R., Y. ZHONG, M. DE ANDRADE, P. R. CLEMENS, R. G. FENWICK, and C. T. CASKEY. 1994. Linkage disequilibria among $(CA)_n$ polymorphisms in the human dystrophin gene and their implications in carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophies. Genomics **21**:567–570.

DERMITZAKIS, E., A. G. CLARK, C. BATARGIAS, A. MARGOULAS, and E. ZOUROS. 1998. Negative covariance suggests mutation bias in a two-locus microsatellite system in the fish *Sparus aurata*. Genetics **150**:1567–1575.

DIB, C., S. FAURÉ, C. FIZAMES et al. (14 co-authors). 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380**:152–154.

DIETRICH, W. F., J. C. MILLER, R. G. STEEN et al. (14 co-authors). 1994. A genetic map of the mousse with 4,006 simple sequence length polymorphism. Nat. Genet. **7**:220–255.

DIRIENZO, A., A. C. PETERSON, J. C. GARZA, A. M. VALDES, M. SLATKIN, and N. B. FREIMER. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA **91**:3166–3170.

ESTOUP, A., and B. ANGERS. 1998. Microsatellites and minisatellites for molecular ecology. Pp. 55–86 *in* G. CARVALHO, ed. Advances in molecular ecology. IOS Press.

ESTOUP, A., and J.-M. CORNUET. 1999. Microsatellite evolution: inferences from population data. Pp. 50–65 *in* D. B. GOLDSTEIN and C. SCHLÖTTERER, eds. Microsatellites: evolution and applications. Oxford University Press, Oxford.

ESTOUP, A., L. GARNERY, M. SOLIGNAC, and J.-M. CORNUET. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140**:679–695.

ESTOUP, A., F. ROUSSET, Y. MICHALAKIS, J.-M. CORNUET, M. ADRIAMANGA, and R. GUYOMARD. 1998. Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Mol. Ecol. **7**:339–353.

ESTOUP, A., M. SOLIGNAC, M. HARRY, and J.-M. CORNUET. 1993. Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res. **21**:1427–1431.

FELDMAN, M. W., A. BERGMAN, D. D. POLLOCK, and D. B. GOLDSTEIN. 1997. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. Genetics **145**:207–216.

GARZA, J. C., M. SLATKIN, and N. B. FREIMER. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol. Biol. Evol. **12**:594–603.

GIUFFRA, E., R. GUYOMARD, and G. FORNERIS. 1996. Phylogenetic relationships and introgression pattern between incipient parapatric species of Italian brown trout (*Salmo trutta* L. complex). Mol. Ecol. **5**:207–220.

GOLDSTEIN, D. B., A. R. LINARES, M. W. FELDMAN, and L. L. CAVALLI-SFORZA. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92**:6723–6727.

GOLDSTEIN, D. B., and D. D. POLLOCK. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J. Hered. **88**:335–342.

GOOSTREY, A., D. N. CARSS, L. R. NOBLE, and S. B. PIERTNEY. 1998. Population introgression and differentiation in the great cormorant *Phalacrocorax carbo* in Europe. Mol. Ecol. **7**:329–338.

GOTTELLI, D., C. SILLERO-ZUBIRI, G. D. APPLEBAUM, M. S. ROY, D. J. GIRMAN, J. GARCIA-MORENO, E. A. OSTRANDER,

and R. K. WAYNE. 1994. Molecular genetics of the most endangered canid: the Ethiopian wolf *Canis simensis.* Mol. Ecol. **3**:301–312.

GYLLENSTEN, U., R. F. LEARY, F. W. ALLENDORF, and A. C. WILSON. 1985. Introgression between two cutthroat subspecies with substantial karyotypic, nuclear and mitochondrial genomic divergence. Genetics **111**:905–915.

HANTULA, J., M. DUSABENYAGASANI, and R. C. HAMELIN. 1996. Random amplified microsatellites (RAMS): a novel method for characterizing genetic variation within fungi. Eur. J. For. Pathol. **26**:159–166.

HARE, M. P., and J. C. AVISE. 1996. Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). Evolution **50**:2305–2315.

HARRISON, R. G. 1993. Hybrid zones and evolutionary process. Oxford University Press, Oxford.

KARL, A. S., and J. C. AVISE. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science **256**:100–102.

KIMURA, M., and T. OHTA. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc. Natl. Acad. Sci. USA **75**:2868–2872.

LEVINSON, G., and G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

LOBO, J. A., M. A. DEL LAMA, and M. A. MESTRINER. 1989. Population differentiation and racial admixture in the Africanized honey bee (*Apis mellifera* L.). Evolution **43**:794–802.

MACHUGH, D. E., M. D. SHRIVER, R. T. LOFTUS, P. CUNNINGHAM, and D. G. BRADLEY. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (*Bos taurus* and *Bos indicus*). Genetics **146**:1071–1086.

MIRSKY, A. E., and H. RISS. 1951. The deoxyribonucleic acid content of animal cells and its evolutionary significance. J. Genet. Physiol. **34**:451–462.

NAUTA, M. J., and F. J. WEISSING. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143**:1021–1032.

PENA, S. D. J., K. T. DE SPUZA, M. DE ANDRADE, and R. CHAKRABORTY. 1994. Allelic associations of two polymorphic microsatellites in intro 40 of the human von Willebrand factor gene. Proc. Natl. Acad. Sci. USA **91**:723–727.

POLLOCK, D. D., A. BERGMAN, M. W. FELDMAN, and D. B. GOLDSTEIN. 1998. Microsatellite behavior with range constraints: parameter estimation and improved distances for use in phylogenetic reconstruction. Theor. Popul. Biol. **53**: 265–271.

POTEAUX, C., F. BONHOMME, and P. BERREBI. 1998. Differences between nuclear and mitochondrial introgressions of brown trout populations from a restocked main river and its unstocked tributary. Biol. J. Linn. Soc. **63**:379–392.

PRIMMER, C. R., N. SAINO, A. P. MOLLER, and H. ELLEGREN. 1998. Unraveling the processes of microsatellite evolution through analysis of germline mutations in barn swallows, *Hirundo rustica.* Mol. Biol. Evol. **15**:1047–1054.

RINDERER, T. E., J. A. STELZER, B. P. OLDROYD, S. M. BUCO, and W. L. RUBINK. 1991. Hybridization between European and Africanized honey bees in the Neotropical Yucatan Peninsula. Science **253**:309–311.

ROY, M. S., E. GEFFEN, D. SMITH, E. A. OSTRANDER, and R. K. WAYNE. 1994. Pattern of differentiation and hybridization in North American wolflike canid, revealed by analysis of microsatellite loci. Mol. Biol. Evol. **11**:553–570.

SCHLÖTTERER, C., and D. TAUTZ. 1992. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**:211–215.

SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141**:413–429.

SOKAL, R. R., and F. J. ROHLF. 1995. Biometry. 3rd edition. Freeman, New York.

SPRENT, P. 1989. Applied nonparametric statistical methods. Chapman and Hall, London.

THOREN, P., R. PAXTON, and A. ESTOUP. 1995. Unusually high frequency of $(CT)_n$ and $(GT)_n$ microsatellites in a yellowjacket wasp, *Vespula rufa* L. (Hymenoptera: Vespidae). Insect Mol. Biol. **4**:141–148.

WEBER, J. L., and C. WONG. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. **2**:1123–1128.

ROSS CROZIER, reviewing editor

Accepted March 9, 1999