Mutation and Evolution of Microsatellite Loci in Neurospora

Jeremy R. Dettman[1] and John W. Taylor

Department of Plant and Microbial Biology, University of California, Berkeley, California 94720

Manuscript received March 26, 2004
Accepted for publication July 14, 2004

ABSTRACT

The patterns of mutation and evolution at 13 microsatellite loci were studied in the filamentous fungal genus Neurospora. First, a detailed investigation was performed on five microsatellite loci by sequencing each microsatellite, together with its nonrepetitive flanking regions, from a set of 147 individuals from eight species of Neurospora. To elucidate the genealogical relationships among microsatellite alleles, repeat number was mapped onto trees constructed from flanking-sequence data. This approach allowed the potentially convergent microsatellite mutations to be placed in the evolutionary context of the less rapidly evolving flanking regions, revealing the complexities of the mutational processes that have generated the allelic diversity conventionally assessed in population genetic studies. In addition to changes in repeat number, frequent substitution mutations within the microsatellites were detected, as were substitutions and insertion/deletions within the flanking regions. By comparing microsatellite and flanking-sequence divergence, clear evidence of interspecific allele length homoplasy and microsatellite mutational saturation was observed, suggesting that these loci are not appropriate for inferring phylogenetic relationships among species. In contrast, little evidence of intraspecific mutational saturation was observed, confirming the utility of these loci for population-level analyses. Frequency distributions of alleles within species were generally consistent with the stepwise mutational model. By comparing variation within species at the microsatellites and the flanking-sequence, estimated microsatellite mutation rates were ~2500 times greater than mutation rates of flanking DNA and were consistent with estimates from yeast and fruit flies. A positive relationship between repeat number and variance in repeat number was significant across three genealogical depths, suggesting that longer microsatellite alleles are more mutable than shorter alleles. To test if the observed patterns of microsatellite variation and mutation could be generalized, an additional eight microsatellite loci were characterized and sequenced from a subset of the same Neurospora individuals.

MICROSATELLITES are composed of tandemly repeated, simple DNA sequence motifs of as many as six nucleotides in length. These loci are commonly found throughout both prokaryotic and eukaryotic genomes and typically are highly polymorphic within species and populations. In addition, these codominant genetic markers are relatively easy to score and have high reproducibility and specificity. As such, microsatellites have become one of the most popular classes of molecular markers and are commonly employed to investigate the population genetics of a diverse range of organisms (BRUFORD and WAYNE 1993; GOLDSTEIN and SCHLÖTTERER 1999). Although fewer studies have addressed the evolutionary dynamics and mutational processes of microsatellite loci, an understanding of these topics is beginning to develop (see ESTOUP and CORNUET 1999; SCHLÖTTERER 2000).

This study of microsatellite evolution in the model microbial eukaryote Neurospora had three main objectives:

1. To describe the patterns of mutation and evolution at multiple microsatellite loci: Are these patterns similar across loci and across species? Are these patterns consistent with accepted theories of microsatellite mutation and evolution?
2. To determine if microsatellites are appropriate genetic markers for inferring the phylogenetic relationships among species of Neurospora: At what level of phylogenetic divergence does microsatellite mutational saturation occur? Are the amounts of allelic homoplasy large enough to render interspecific comparisons meaningless?
3. To determine if microsatellites would be useful for answering population genetic questions: What are the relative rates of microsatellite mutation, and how variable are these genetic markers within a single species? What factors can be used to predict microsatellite variability?

Slipped-strand mispairing during DNA replication (LEVINSON and GUTMAN 1987) is considered the pre-

[1]*Corresponding author:* Department of Botany, University of Toronto, 3359 Mississauga Rd., Mississauga, ON L5L 1C6, Canada. E-mail: jdettman@utm.utoronto.ca

dominant mutational mechanism for microsatellites (SCHLÖTTERER and TAUTZ 1992; STRAND *et al.* 1993). During replication, the nascent strand may dissociate from the template strand and, due to the repetitive nature of the microsatellite DNA, the strands may reanneal incorrectly, or "out-of-register." The pairing of misaligned repeats causes one to several repeats to loop out from either the nascent or the template strand. This slipped-strand mispairing results in the nascent strand having a different number of repeats from the template strand once DNA replication is complete. This mutation process allows the same microsatellite allele to arise multiple times, thus generating size homoplasy. A number of mutational models that account for this homoplasy have been proposed, with the stepwise mutational model (SMM; KIMURA and OHTA 1978) and the two-phase model (TPM; DI RIENZO *et al.* 1994) being the most common. The SMM assumes all mutational events involve a change in a single repeat only, while the TPM allows a proportion of mutations to involve changes greater than single repeats. On the other hand, the infinite-allele model (IAM; KIMURA and CROW 1964) does not allow for homoplasy and assumes that every mutation results in the creation of a new allele. Determining which mutational model is most appropriate is important because microsatellite-specific genetic distances rely on the underlying assumptions of the chosen model (TAKEZAKI and NEI 1996).

Several outbreeding species of Neurospora have been recognized both phylogenetically (DETTMAN *et al.* 2003a) and in terms of reproductive isolation (TURNER *et al.* 2001; DETTMAN *et al.* 2003b); however, many of the basal relationships among species remain unresolved. A number of these species are included in ongoing studies of comparative genomics (GILBERT *et al.* 2003) and the genetics of reproductive isolation (TURNER and TAYLOR 2003), underscoring the need for a reliable phylogeny of the genus. Molecular data suggest that the group underwent a rapid species radiation (NATVIG and MAY 1996; SKUPSKI *et al.* 1997; DETTMAN *et al.* 2003a). Protein-coding genes, and even noncoding genomic regions, appear to have evolved too slowly to record the full speciation history (*i.e.*, few mutations arose and became fixed between successive speciation events). Because microsatellites mutate at such high rates, they may possess enough variation to clarify interspecific relationships. However, the consequence of hypermutability combined with homoplasy is the complete loss of phylogenetic signal over long periods of evolution. Thus, microsatellites that are conserved across diverse taxonomic groups (*e.g.*, SCHLÖTTERER *et al.* 1991; FITZSIMMONS *et al.* 1995; ZARDOYA *et al.* 1996) may not be appropriate for inferring the phylogenetic relationships among the groups. Before constructing a phylogeny of Neurospora species using microsatellites, we assessed whether or not we could place confidence in the resulting tree.

A common approach to the study of the evolutionary dynamics and mutational processes of microsatellites has been to infer past mutational events from contemporary allele frequency distributions within populations (*e.g.*, SHRIVER *et al.* 1993; VALDES *et al.* 1993; DI RIENZO *et al.* 1994; ESTOUP *et al.* 1995a,b; GARZA *et al.* 1995; NIELSEN and PALSBØLL 1999). This method typically requires electromorph typing of a large number of loci from hundreds of individuals. A different approach is to directly sequence the microsatellite and flanking regions to reveal mutations that simple electromorph typing cannot detect (*e.g.*, ESTOUP *et al.* 1995b; JIN *et al.* 1996; ANGERS and BERNATCHEZ 1997; GRIMALDI and CROUAU-ROY 1997; ORTÍ *et al.* 1997; PRIMMER and ELLEGREN 1998; COLSON and GOLDSTEIN 1999; FISHER *et al.* 2000; HARR *et al.* 2000; MAKOVA *et al.* 2000; ZHU *et al.* 2000; BLANKENSHIP *et al.* 2002). This approach is more laborious than the former so sequencing usually has been limited to a smaller number of individuals and/or loci.

We studied the evolutionary dynamics of 13 microsatellite loci in the filamentous fungal genus Neurospora by taking an approach that was a combination of those mentioned above. First, a detailed investigation was performed on five microsatellite loci by sequencing the repeat arrays, along with ∼500 nucleotides of their flanking regions, from a set of 147 individuals from eight phylogenetic species of Neurospora. To elucidate the genealogical relationships among microsatellite alleles, repeat number was mapped onto trees constructed from flanking-sequence data. This approach allowed the microsatellite mutations to be placed in the evolutionary context of the less rapidly evolving, nonrepetitive, flanking DNA. In addition to assessing variation and testing microsatellite mutation models, three types of novel analyses were performed. By comparing microsatellite and flanking-sequence genetic distances, the phylogenetic depth at which microsatellite mutational saturation occurred could be estimated. Variation within species at the microsatellites and the linked flanking regions were used to estimate the relative rates of mutation for the two DNA types. The hierarchical sampling scheme of this study provided a unique opportunity to investigate the evolution of microsatellites across multiple genealogical depths. To further investigate the evolution of microsatellite repeat arrays, an additional eight microsatellite loci were characterized and sequenced from a subset of 29 individuals containing representatives from each of the same eight Neurospora species. A total of 921 microsatellite alleles were directly sequenced, which represents the largest effort put toward microsatellite sequencing to date.

MATERIALS AND METHODS

Selection of individuals: The 147 individuals used in this study were described by DETTMAN *et al.* (2003a) and are listed

in the supplemental data. The entire sample included individuals from eight phylogenetic species of Neurospora: *N. crassa* (n = 33, 3, and 9 from NcA, NcB, and NcC subgroups, respectively); *N. intermedia* (n = 52 and 10 from NiA and NiB subgroups, respectively, and 6 from basal lineages); *N. sitophila* (n = 7); *N. tetrasperma* (n = 4); phylogenetic species 1 (n = 3); phylogenetic species 2 (n = 7); phylogenetic species 3 (n = 5); and *N. discreta* (n = 8). Some species are referred to simply as "phylogenetic species" because they have not yet been formally named (see DETTMAN *et al.* 2003a,b).

Microsatellite loci: Sequences of four nonrepetitive DNA regions, within which five microsatellite loci were located, were reported in DETTMAN *et al.* (2003a; GenBank accessions AY225899–AY225949; TREEBASE accessions S950 and M1574). Here we report the sequences of the repeat arrays for these five microsatellite loci (DMG-CA, TMI-AGC, TML-ATC, QMA-CTTTT, and QMA-TGG) from 147 individuals (see supplemental data at http://www.genetics.org/supplemental/).

New primer sets were designed for an additional eight microsatellite loci from five genomic regions that were unlinked to the previous four genomic regions. Various microsatellite repeat motifs (tandemly repeated 15 times) were used to search the Broad Institute *N. crassa* Database (http://www.broad.mit.edu/annotation/fungi/neurospora/). Iterative rounds of primer design were performed until a 400- to 600-bp fragment (microsatellite plus flanking regions) could be amplified from the majority of species. These eight new microsatellites were sequenced from a subset of 29 individuals from eight species (see supplemental data).

On the basis of open reading frame predictions provided by the genome database and the form and distribution of variation among the sequences from different individuals, three of the microsatellites (TMI-AGC, T6L-AGC, and T7L-ACA) may be located within coding DNA. However, the existence of these predicted genes and hypothetical proteins has not been verified.

PCR amplification conditions and sequencing protocols were the same as described in DETTMAN *et al.* (2003a) except for locus-specific annealing temperatures (Table 1). For all loci, microsatellite repeat number was determined directly from the sequence data and included both perfect and imperfect repeats. When substitutions or noninteger length mutations made counting repeats difficult, repeat numbers were calculated by dividing array length by motif length and then rounding to nearest integer.

Data analyses: The program Microsatellite Analyser (MSA, version 3.0; DIERINGER and SCHLÖTTERER 2003) was used to calculate descriptive statistics and microsatellite genetic distances among groups. Neighbor-joining (SAITOU and NEI 1987) trees were constructed from genetic distances by importing the distance matrices into PAUP (version 4b10; SWOFFORD 2001). Branch support was assessed by performing 100 bootstrap replications using MSA. The topologies of the microsatellite trees were compared to that of the flanking-sequence tree from DETTMAN *et al.* (2003a) using Shimodaira-Hasegawa tests (SHIMODAIRA and HASEGAWA 1999; resampling estimated log likelihoods from 1000 bootstraps). For these tests, the reference data set contained combined sequence data from the DMG, TMI, TML, and QMA genomic regions for a single individual from each of the eight species, and maximum-likelihood parameter values were estimated using ModelTest (version 3.06; POSADA and CRANDALL 1998).

The program JMPin (version 3.2.6; SAS INSTITUTE 1999) was used for analysis of covariance (ANCOVA), correlation, and regression analyses of both nontransformed and log-transformed data. Unless otherwise noted, the results and conclusions of analyses did not differ between the nontransformed and log-transformed data. All flanking-sequence genetic distances reported in this study are Kimura two-parameter distances, as calculated by PAUP. The program SITES (HEY and WAKELEY 1997) was used to calculate numbers of segregating sites and to estimate the population mutation parameter θ from sequence data. For haploid organisms such as Neurospora, θ equals two times the effective population size times the mutation rate ($\theta = 2N\mu$), assuming neutrality and mutation-drift equilibrium.

Mutational saturation of a microsatellite was assessed as follows: For each possible pairwise comparison of individuals, we calculated the number of substitution differences between flanking sequences and the commonly employed microsatellite genetic distance $(\delta\mu)^2$ (GOLDSTEIN *et al.* 1995a) between microsatellite alleles. When comparing individuals, $(\delta\mu)^2$ is functionally equivalent to the measure D_1, which is the squared difference in repeat number (GOLDSTEIN *et al.* 1995b). For each microsatellite, we then plotted the mean $(\delta\mu)^2$ distances against the number of substitutions in flanking sequence. Because the expectation of $(\delta\mu)^2$ is linear with respect to time (GOLDSTEIN *et al.* 1995a), mutational saturation was estimated to occur at the point where mean $(\delta\mu)^2$ no longer increased with the increasing number of flanking-sequence substitutions. To account for differences in flanking-sequence length, these comparisons were repeated using genetic distances between the flanking sequences (which were binned), and the same results were obtained. Using difference in repeat number between microsatellite alleles rather than $(\delta\mu)^2$ also did not affect the results. These calculations assumed that the microsatellites and their physically adjacent flanking regions were completely linked. Correlations between microsatellite distances and flanking-sequence genetic distances for each microsatellite, and correlations among flanking-sequence genetic distances among genomic regions, were calculated from all possible pairwise comparisons of individuals.

Evolution under the SMM commonly results in a unimodal, symmetrical distribution of repeat alleles around the mean, whereas evolution under the TPM is more likely to result in irregular, multimodal distributions. However, multimodal distributions alone cannot justify the rejection of the SMM because evolution under SMM can produce these distributions by chance (VALDES *et al.* 1993). Therefore, the fit of microsatellite mutation models to *N. crassa* and *N. intermedia* allele-frequency distributions was tested using a likelihood approach implemented by MISAT (NIELSEN 1997). The fitting of models and θ estimation were restricted to intraspecific subgroups (NcA and NiA) to avoid pooling genetically differentiated groups. Markov chains were run for 100,000 and 1,000,000 generations for the SMM and TPM, respectively. For the TPM, the proportion of multistep mutations was allowed to vary between 0.00001 and 0.5. Microsatellite θ value estimation and likelihood-ratio tests were performed as described by NIELSEN (1997). These calculations, along with θ comparisons between microsatellites and flanking regions, rest on the assumption that the populations have reached mutation-drift equilibrium.

RESULTS

Overall variation: The evolution of 13 microsatellite loci from Neurospora was investigated in this study. These microsatellites are embedded in nine different genomic regions, each from a different chromosome arm or linkage group (Table 1). To distinguish the microsatellite from its genomic region, the name of the microsatellite proper includes a dash followed by a repeat motif. For example, TMI-AGC is a microsatellite with an AGC repeat motif and is embedded within the

TABLE 1

Information on the 13 microsatellites characterized in this study

Microsatellite locus	Linkage group	BINCD[a] contig/supercontig	Location of center of microsatellite[b]	Primers (5′–3′)	Annealing temperature
DMG-CA	IVR	3.379/29	72292(−)	See Dettman *et al.* (2003a)	
TMI-AGC	VR	3.51/3	29355	See Dettman *et al.* (2003a)	
TML-ATC	VL[c]	3.555/61	40717(−)	See Dettman *et al.* (2003a)	
QMA-CTTTT	IIR	3.307/20	41535	See Dettman *et al.* (2003a)	
QMA-TGG	IIR	3.307/20	41641	See Dettman *et al.* (2003a)	
T1R-AT	IR	3.146/8	73950	Forward is T1R-F1: TGTACTCGGATGACGTGCTA	53.4
				Reverse is T1R-R1: AAGGCGGTTTCTGTTATCAA	
T1R-TAG	IR	3.146/8	73998	Same as T1R-AT	
T1R-AGC	IR	3.146/8	74237	Same as T1R-AT	
T2L-ATC	IIL	3.181/10	3629	Forward is T2L-F1: CACCACCATCGTCGTCACAC	57.5
				Reverse is T2L-R1: GGACGGAGTCATCCAGCATC	
				Forward is T2L-F2: CACCACCATCGTCGTCACACG	58.9
				Reverse is T2L-R2: GAGGGACGGAGTCATCCAGCA	
T3L-AGC	IIIL	3.319/22	11096	Forward is T3L-F1: CTAGGTAGTGCCAGGAGTGC	55.8
				Reverse is T3L-R1: GACATCCTCGCCATAGTACC	
				Reverse is T3L-R2: CATCGACCGAAAGACCTTTC	
T6L-AGC	VIL	3.262/16	24711(−)	Forward is T6L-F1: CGTCGTAGTAAGATTCCGTA	55.4
				Reverse is T6L-R1: ACTCTTTGAGACCGCTTTAC	
T7L-ACA	VIIL	3.350/25	3709(−)	Forward is T7L-F1: CTCGCTGACACTTCGGCATC	57.5
				Forward is T7L-F2: CATCGCACGAGCCAATTTTA	
				Reverse is T7L-R1: CGGCCAGAACGGTCACTGTA	
T7L-CGA	VIIL	3.350/25	3945(−)	Same as T7L-ACA	

[a] BINCD, Broad Institute *Neurospora crassa* Database.
[b] (−) indicates that the sequence is from the minus strand.
[c] As predicted by the Munich Information Centre for Protein Sequences *Neurospora crassa* Database.

∼500-bp stretch of PCR-amplified sequence called the TMI genomic region. Similarly, QMA-CTTTT and QMA-TGG are two linked microsatellites that are both embedded within the QMA genomic region.

Five microsatellite loci, located on four different chromosome arms, were sequenced from 147 individuals from eight Neurospora species (Table 2). The repeat numbers of the microsatellite alleles are listed in the supplemental data, along with the full sequences of alleles. The patterns of variation were quite different among these microsatellite loci. Each locus possessed between 11 and 21 different repeat alleles, and there was an eightfold difference among loci in variance in repeat number (Table 2). The dinucleotide microsatellite DMG-CA was the most variable locus with respect to the number of repeat alleles and variance in repeat number. The frequency distributions of repeat alleles also differed drastically among loci (see supplemental data at http://www.genetics.org/supplemental/). For example, DMG-CA had a wide range of intermediate-frequency repeat alleles, while QMA-TGG had one main high-frequency repeat allele.

A significant amount of variation was found in the nonrepetitive regions flanking the microsatellites, even within a single species (Table 2, Figure 1). Previous authors (Brohede and Ellegren 1999) have reported that the variability of nonrepetitive DNA increases with proximity to microsatellites. However, polymorphism plots across the length of the Neurospora sequence alignments provided no evidence for such a trend (data

**Summary of molecular variation observed at the five microsatellite loci investigated in detail,
along with variation observed in the flanking regions**

DNA type	Microsatellite locus				
	DMG-CA	TMI-AGC	TML-ATC	QMA-CTTTT	QMA-TGG
Microsatellite					
Sample size	147	147	146	130	124
No. of repeat alleles[a]	21	16	11	16	13
No. of unique alleles[b]	46	20	38	47	23
Repeat no. (min, max)	4, 37	4, 20	5, 15	5, 43	2, 17
Mean repeat number	12.61	9.27	7.59	9.52	5.89
Variance in repeat no.	43.16	10.32	6.62	25.03	5.31
Nonrepetitive flanking region					
Sample size	147	147	147	145	145
Mean no. of included nucleotides	378	445	538	482	482
No. of flanking-sequence variants	61	66	95	85	85
Mean genetic distance	0.0081	0.0371	0.0294	0.0839	0.0839

[a] Alleles with different numbers of tandem repeats.
[b] Alleles with different DNA sequence.

not shown). The number of unique flanking-sequence variants increased with the length of included flanking sequence per region (Table 2). Interestingly, the relative levels of variation per microsatellite did not correspond with the relative levels of variation in the linked, nonrepetitive flanking regions. The QMA flanking region was the most variable, but QMA-TGG had the lowest variance in repeat number, whereas the DMG flanking region was the least variable, but DMG-CA had the highest variance in repeat number (Table 2). Thus, highly variable microsatellite loci were not found only in highly variable genomic regions.

Eight additional microsatellite loci, located on five different chromosome arms, were sequenced from the 29 individuals that were representatives of the eight Neurospora species (see supplemental data). Each locus possessed between 6 and 11 different repeat alleles, and there was a 35-fold difference among loci in variance in repeat number (Table 3). Even for two microsatellites located within the same genomic region, there could be a 13-fold difference in variability (*e.g.*, T7L-ACA and T7L-CGA).

The following sections present analyses that were performed primarily on data from the five intensively sampled microsatellite loci. Most of the same analyses were performed on data from the additional eight loci, and results typically were similar. Analyses of the five-locus data set had more power because the sample size per locus was significantly larger. For the sake of brevity, analyses of the eight-locus data set are discussed only if they provide novel information.

Mutational saturation and homoplasy: For each of the four genomic regions investigated in detail, the genealogies presented in DETTMAN *et al.* (2003a) were pruned to retain only a single representative of each flanking-

sequence variant (Figure 1). The microsatellite repeat alleles that were associated with each flanking-sequence variant then were mapped onto the trees. As predicted by the fact that microsatellites are hypermutable, many flanking-sequence variants were associated with multiple repeat alleles. However, many repeat alleles also were shared by highly divergent flanking-sequence variants. Although identical in state, it was very unlikely that all copies of the same repeat allele were identical by descent, especially when shared among well-differentiated genealogical lineages (Figure 1).

Microsatellite mutation following the SMM tends to create derived alleles that have a similar number of repeats as the ancestral allele. Difference in repeat number therefore holds some information on the recency of common ancestry, providing the basis for SMM-based genetic distances (GOLDSTEIN *et al.* 1995a,b). Because the number of substitution differences among nonrepetitive DNA also is proportional to recency of common ancestry, a correlation between microsatellite distance and flanking-sequence genetic distance is expected, at least until the divergence becomes so great that mutational saturation begins to degrade the phylogenetic signal. Despite the physical linkage of the microsatellites and their respective flanking regions, there was a poor correlation between flanking-sequence genetic distance and $(\delta\mu)^2$, a microsatellite-specific distance (mean R^2 for five microsatellite-by-flanking region comparisons = 0.02). On the other hand, flanking-sequence genetic distances were well correlated among the four genomic regions (mean R^2 for six flanking region-by-flanking region comparisons = 0.64), suggesting that the different nonrepetitive regions have been accumulating mutations in a more or less linear fashion without significant substitutional saturation. The poor correlation between
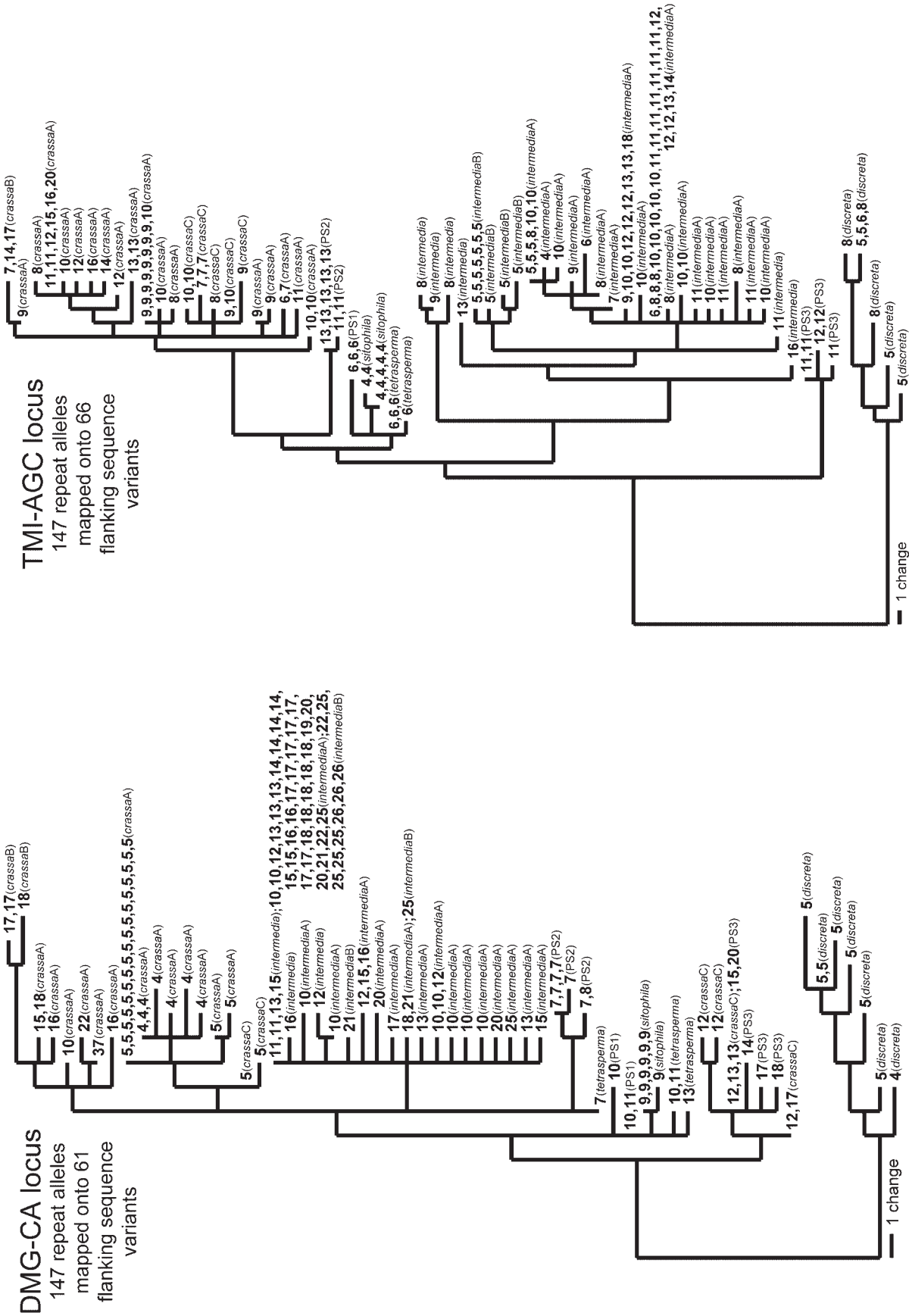
FIGURE 1.—Microsatellite repeat alleles mapped onto maximum-parsimony phylograms constructed from the nonrepetitive flanking DNA. For flanking-sequence variants that were sampled multiple times, multiple microsatellite alleles are listed. Names in parentheses following repeat numbers indicate the species (and subgroup) from which the allele was sampled. See supplemental data for complete sequence data for all microsatellite alleles.
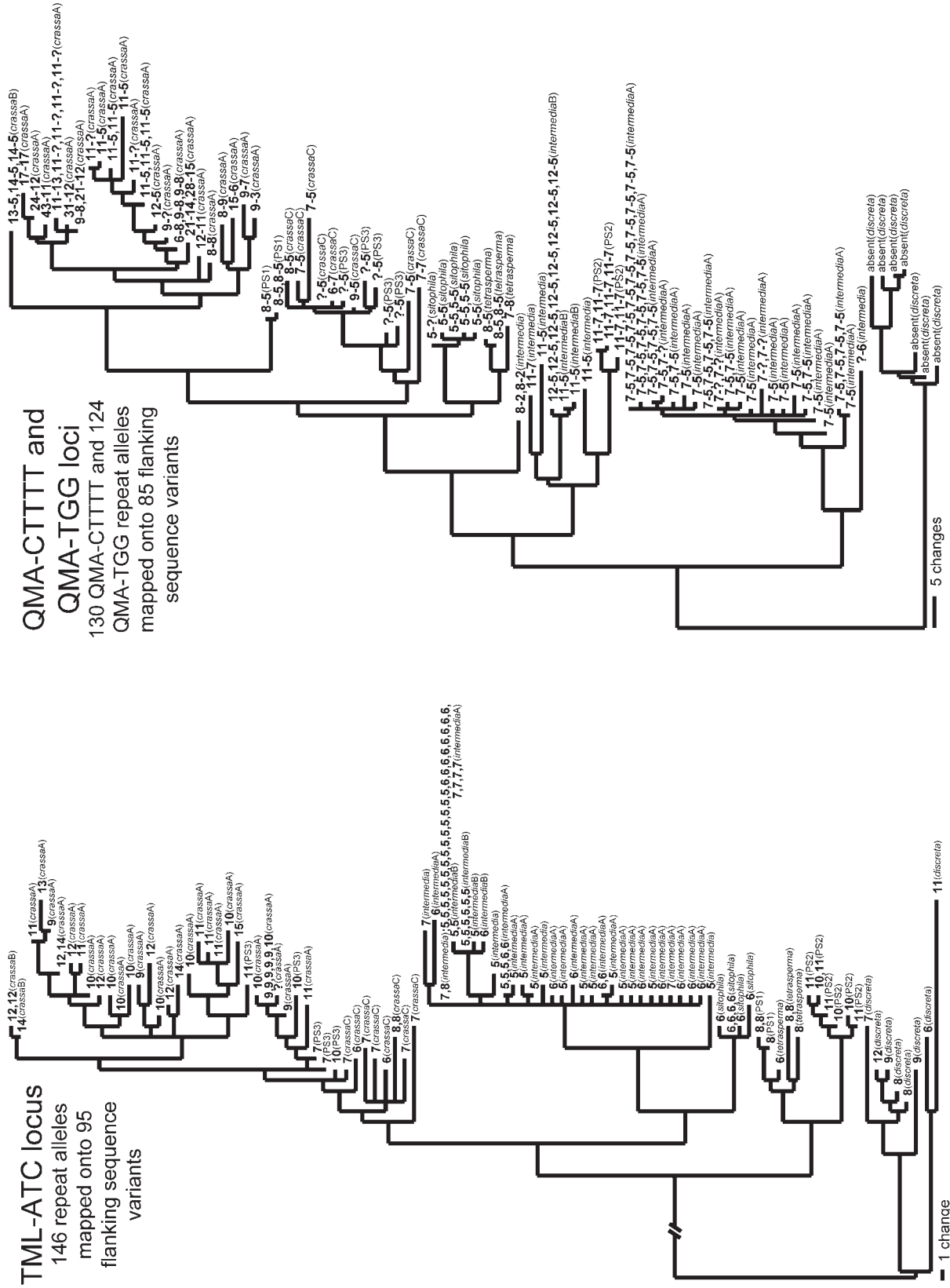
FIGURE 1.—*Continued.*

TABLE 3

Summary of molecular variation observed at the additional eight microsatellite loci

	Microsatellite locus							
	T1R-AT	T1R-TAG	T1R-AGC	T2L-ATC	T3L-AGC	T6L-AGC	T7L-ACA	T7L-CGA
Sample size	29	29	29	28	29	29	27	27
No. of repeat alleles[a]	6	10	11	9	9	6	11	9
No. of unique alleles[b]	8	17	21	13	14	9	12	12
Repeat number (min, max)	3, 11	9, 20	5, 19	6, 18	6, 15	4, 9	5, 46	4, 13
Mean repeat no.	5.48	14.55	10.10	7.89	9.62	6.45	13.85	9.15
Variance in repeat no.	6.27	20.64	11.43	10.10	6.18	2.64	93.42	7.18

[a] Alleles with different numbers of tandem repeats.

[b] Alleles with different DNA sequence.

variation in the flanking regions and linked microsatellites was presumably due to mutational saturation and homoplasy of the microsatellites. If this hypothesis is true, the correlation between variation levels of the two DNA types would be expected to increase as comparisons are made over shallower genealogical depths.

At what level of divergence does microsatellite homoplasy begin to erase the history of the alleles? By plotting $(\delta\mu)^2$ against the number of substitutions between the flanking sequences or binned genetic distances to account for differences in flanking-sequence length, mutational saturation was estimated to begin when flanking-sequence genetic distances reached 0.006, 0.012, 0.048, 0.017, and 0.017 for DMG-CA, TMI-AGC, TML-ATC, QMA-CTTTT, and QMA-TGG, respectively (Figure 2, Table 4). Results were similar when flanking-sequence genetic distance or difference in microsatellite repeat number were used. As expected, the most variable of the five microsatellite loci, DMG-CA, was mutationally saturated at the smallest flanking-sequence genetic distance. On average, the ability of these microsatellite loci to represent evolutionary history begins to degrade at 2.0% divergence of flanking regions.

Comparison among species: Four of the five microsatellite loci reached mutational saturation at a flanking-sequence genetic distance that was less than the mean genetic distance among the seven ingroup species, *i.e.*, excluding *N. discreta* (Table 4). Averaged across loci, the flanking-sequence divergence among ingroup species was 3.3%, far beyond the 2.0% divergence level at which the microsatellites were shown to saturate. In addition, the range of microsatellite repeat number commonly overlapped among species. The percentage of repeat alleles shared by at least two species was 66.7, 75.0, 63.6, 25.0, and 30.8% for DMG-CA, TMI-AGC, TML-ATC, QMA-CTTTT, and QMA-TGG, respectively. This saturation and excessive homoplasy suggested that interspecific phylogenies based on microsatellite data may be misleading.

Microsatellite genetic distances among the eight Neurospora species were calculated. The tree constructed from $(\delta\mu)^2$, an SMM-based genetic distance, was topologically incongruent with the species phylogeny inferred from the flanking-sequence data (Figure 3; Shimodaira-Hasegawa test, $P = 0.009$). Even when using an IAM-based genetic distance (proportion of shared alleles (POSA); BOWCOCK *et al.* 1994), which ignores difference in repeat number, the tree was topologically incongruent with the flanking-sequence tree (Figure 3, $P = 0.019$). Neither of the two sister-group relationships supported by the flanking-sequence data (PS3-*N. crassa* and PS2-*N. intermedia*) appeared in the microsatellite-based trees. When data from the eight additional microsatellite loci were combined with these five loci, the resulting trees (not shown) were even less resolved and still topologically incongruent with the sequence-based tree.

Variance in repeat number: Mean repeat number did not differ significantly among species, but species did differ significantly in terms of maximum repeat number and variance in repeat number (Mann-Whitney U-tests, $P < 0.05$). These values tended to be greatest in *N. crassa* and *N. intermedia* (Table 5); however, these two species also had the largest sample sizes. To determine which variables were the best predictors of variance in repeat number, we modeled locus, species, sample size, maximum, and mean repeat number using ANCOVA. When all of these variables were accounted for, the only ones that were significant predictors of variance in repeat number were maximum and mean repeat number (both $P < 0.0001$).

Some genealogical lineages were fixed for one repeat allele or possessed only a few, whereas other lineages possessed a wide range of repeat alleles (Figure 1). Although shallower clades tended to be less variable, similar-depth clades could differ greatly in variability, and clades could be fixed for a single repeat allele, regardless of clade depth. When a clade was fixed for a single allele, that allele tended to have a low number of repeats, and when a clade was highly variable, the alleles tended to have higher numbers of repeats (Figure 1), consistent with ANCOVA results.
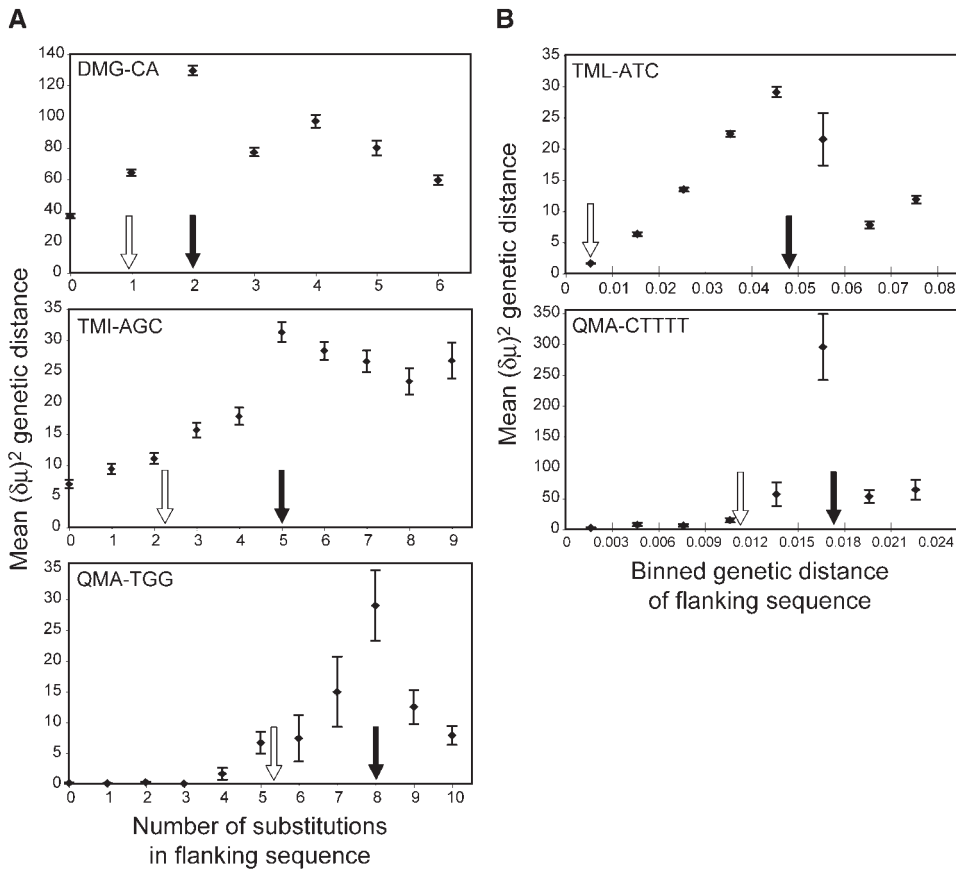
A

B

FIGURE 2.—For each microsatellite locus, mean $(\delta\mu)^2$ genetic distance was plotted against (A) number of substitutions and (B) binned genetic distance between the flanking sequences. Both methods yielded similar results, so plots are shown for only one method per locus. Vertical bars indicate standard error. Solid arrows indicate estimated points of microsatellite mutational saturation. Open arrows indicate the mean number of substitutions or genetic distance between flanking sequences for intraspecific comparisons only, all of which are less than the estimated points of saturation.

To investigate the robustness of the positive relationship between variance and repeat number, we plotted the variance in repeat number against maximum and mean repeat number for three genealogical depths: (1) flanking-sequence variants that were sampled four or more times ($n = 32$ variants): these comparisons were restricted to the shallowest genealogical depth possible to decrease the likelihood of homoplasy; (2) intraspecific values for the five intensively sampled loci ($n = 37$ species-by-locus combinations); and (3) the entire sample of 29 individuals from eight species ($n = 13$ loci): by using the variance of the entire sample, possible species-specific effects would be subsumed. At all three depths, there was a significant positive correlation between variance in repeat number and both maximum and mean repeat number (all $P < 0.008$). Variance in repeat number was best correlated with maximum repeat number (Figure 4). To estimate the lower threshold for slippage, variance in repeat number was linearly fit to maximum and mean repeat number for the 37 species-by-locus combinations, and the x-intercepts were determined (where variance in repeat number = 0). Using these calculations, only microsatellites with a maximum of ≥ 7.70 repeats or a mean of ≥ 5.08 repeats

TABLE 4

Mutational saturation estimates compared to genetic distances within and among species

	Estimated flanking-sequence genetic distance at which microsatellite mutation saturation occurs	Mean flanking-sequence genetic distance *within* species	Mean flanking-sequence genetic distance *among* ingroup species
DMG-CA	0.006	0.003	0.008
TMI-AGC	0.012	0.005	0.029
TML-ATC	0.048	0.005	0.028
QMA-CTTTT	0.017	0.011	0.067
QMA-TGG	0.017	0.011	0.067
Mean	0.020	0.006[a]	0.033[a]

[a] Mean of the four genomic regions.
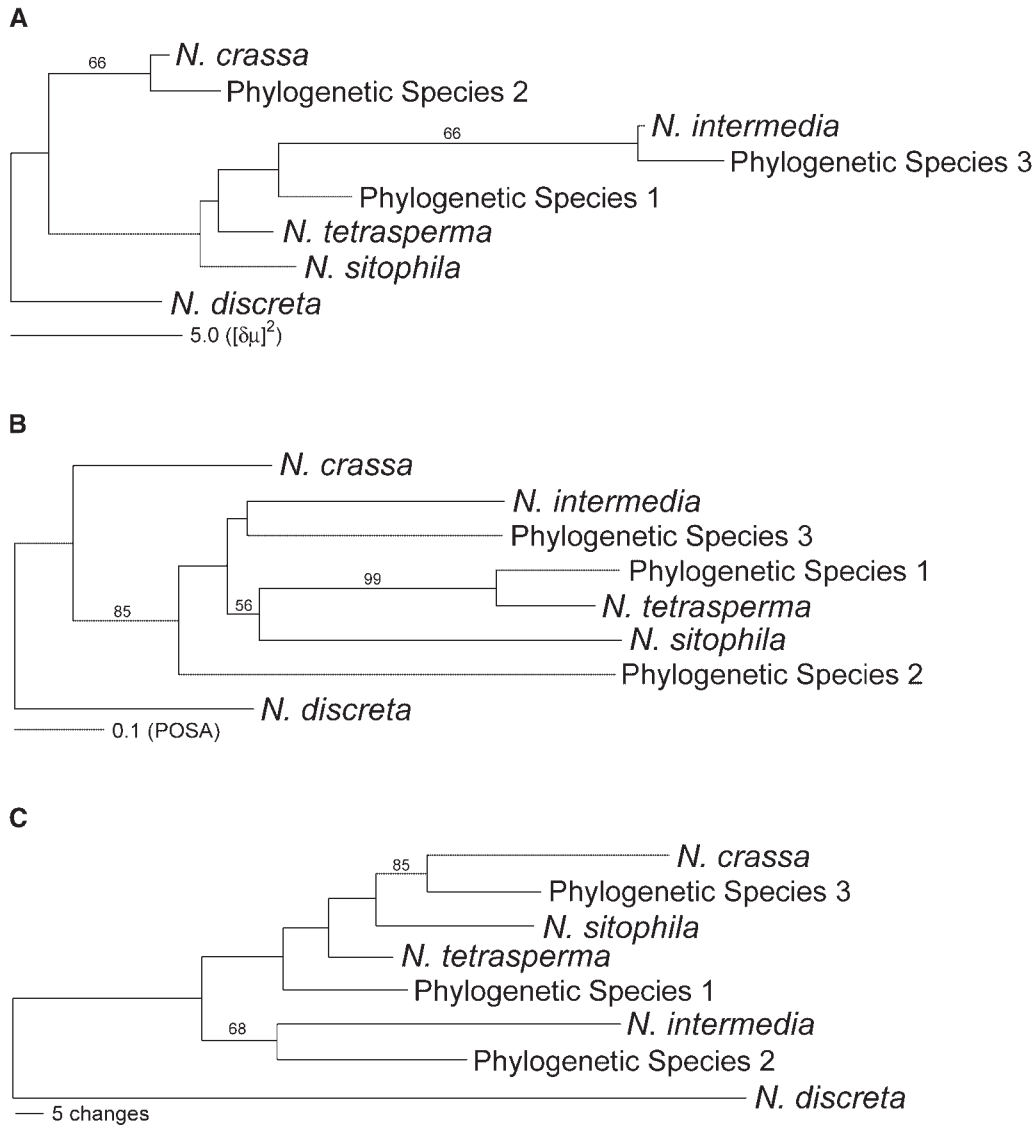
A

B

C

FIGURE 3.—(A) Neighbor-joining tree constructed from SMM-based genetic distances $[(\delta\mu)^2]$ calculated from data from five microsatellite loci. (B) Neighbor-joining tree constructed from IAM-based genetic distances from five microsatellite loci. (C) Approximation of maximum-parsimony tree constructed from flanking-sequence data from four genomic regions (DETTMAN *et al.* 2003a). Numbers near branches indicate bootstrap proportions, if >50.

are expected to be variable within a species. The high prevalence of alleles with five to seven repeats in clades that were fixed for a single allele (Figure 1) was consistent with this threshold.

Substitution mutations within microsatellite repeat arrays: The nine genomic regions were successfully amplified from all eight species of Neurospora, with the exception of T7L from *N. sitophila*. The basic repeat motifs of the 13 microsatellite loci generally were well conserved across species with the exception of QMA-CTTTT and QMA-TGG from *N. discreta,* the outgroup. However, sequence data revealed that substitution mutations within the repeat arrays were common (see supplemental data). For the five loci investigated in detail, the mean number of repeat alleles, or alleles with different numbers of tandem repeats, was 15.4/locus. When repeat number and sequence variation were considered together, the mean number of unique microsatellite alleles increased to 34.8/locus (Table 2). For the eight additional loci, the mean number of repeat alleles was 8.9/locus, but the mean number of unique microsatellite alleles was 13.3/locus (Table 3). Thus, between 55.7% (five loci) and 33.1% (eight loci) of the total allelic variation at microsatellites would have been missed if sequence information had not been obtained.

Several substitution mutations within the repeat arrays were unique and diagnostic for single species. In addition, a number of substitutions were shared among species, many of which were consistent with phylogenetic inferences drawn from the flanking regions (DETTMAN *et al.* 2003a; Figure 3). For example, the PS2-*N. intermedia* sister-group relationship was supported by substitutions in TML-ATC and T7L-ACA microsatellites, while the PS3-*N. crassa* sister-group relationship was supported by substitutions in T1R-TAG and T7L-CGA microsatellites.

The duplication of repeats with substitutions (imperfect repeats) within the same microsatellite allele was observed in DMG-CA, TML-ATC, QMA-CTTTT, QMA-TGG, T1R-TAG, T1R-AGC, 3L-AGC, T7L-ACA, and T7L-CGA (examples in Figure 5). Such examples of substitu-

TABLE 5

Summary of polymorphism at five microsatellites in eight species of Neurospora

Locus	Species	n	Repeat no. (min, max)	Mean repeat no.	Variance in repeat no.
DMG-CA	*N. crassa*	45	4, 37	9.11	46.99
	N. intermedia	68	10, 26	16.71	24.21
	Phylogenetic species 1	3	10, 11	10.33	0.67
	Phylogenetic species 2	7	7, 8	7.14	0.17
	Phylogenetic species 3	5	14, 20	16.80	7.60
	N. sitophila	7	9, 9	9.00	0.00
	N. tetrasperma	4	7, 13	10.25	9.38
	N. discreta	8	4, 5	4.88	0.15
TMI-AGC	*N. crassa*	45	6, 20	10.42	9.05
	N. intermedia	68	4, 18	9.26	8.96
	Phylogenetic species 1	3	6, 6	6.00	0.00
	Phylogenetic species 2	7	11, 13	12.43	1.14
	Phylogenetic species 3	5	11, 12	11.40	0.40
	N. sitophila	7	4, 4	4.00	0.00
	N. tetrasperma	4	6, 6	6.00	0.00
	N. discreta	8	5, 8	6.25	2.58
TML-ATC	*N. crassa*	44	6, 15	10.11	4.96
	N. intermedia	68	5, 8	5.56	0.56
	Phylogenetic species 1	3	8, 8	8.00	0.00
	Phylogenetic species 2	7	10, 11	10.57	0.34
	Phylogenetic species 3	5	7, 11	9.00	4.67
	N. sitophila	7	6, 6	6.00	0.00
	N. tetrasperma	4	6, 8	7.50	1.50
	N. discreta	8	6, 12	8.75	4.58
QMA-CTTTT	*N. crassa*	43	6, 43	12.70	53.73
	N. intermedia	66	7, 12	7.94	3.50
	Phylogenetic species 1	3	8, 8	8.00	0.00
	Phylogenetic species 2	7	11, 11	11.00	0.00
	N. sitophila	7	5, 5	5.00	0.00
	N. tetrasperma	4	7, 8	7.75	0.38
QMA-TGG	*N. crassa*	37	3, 17	7.59	12.48
	N. intermedia	62	2, 7	4.95	0.38
	Phylogenetic species 1	3	5, 5	5.00	0.00
	Phylogenetic species 2	7	7, 7	7.00	0.00
	Phylogenetic species 3	5	5, 5	5.00	0.00
	N. sitophila	6	5, 5	5.00	0.00
	N. tetrasperma	4	5, 8	5.75	3.38

tions that had been multiplied tandemly or along with other repeats without substitutions (perfect repeats) provided evidence that the microsatellites had mutated by slipped-strand mispairing. This explanation was more parsimonious than multiple identical substitutions occurring independently in multiple repeats. Within the same allele, duplicated imperfections that were not adjacent to each other were observed at TML-ATC, QMA-CTTTT, QMA-TGG, T1R-TAG, T1R-AGC, T7L-ACA, and T7L-CGA (examples in Figure 5). Such nonadjacent imperfections were evidence for multistep mutational events, because duplication of imperfect repeats via single-step mutations would create only imperfections that were immediately adjacent.

Intraspecific analyses of *N. crassa* and *N. intermedia*: The following intraspecific analyses were restricted to the two species with sufficient samples sizes, *N. crassa*

and *N. intermedia*, and to the largest subgroup within each species, NcA ($n = 33$) and NiA ($n = 52$), respectively. For each of the five microsatellites investigated in detail, the mean flanking-sequence genetic distance within species was less than the flanking-sequence genetic distance at which microsatellite mutational saturation occurred (Table 4, Figure 2). Because mutational saturation was not a significant problem at intraspecific levels of phylogenetic divergence, these microsatellites are appropriate for use in population genetic studies.

Mutational patterns: To assess the applicability of the SMM within species, frequency distributions of repeat alleles were fit to SMM predictions. The frequency distributions of repeat alleles for NcA and NiA differed among the five loci (Figure 6). At DMG-CA and TMI-AGC, both NcA and NiA were highly variable and had overlapping repeat allele ranges in contrast to the non-
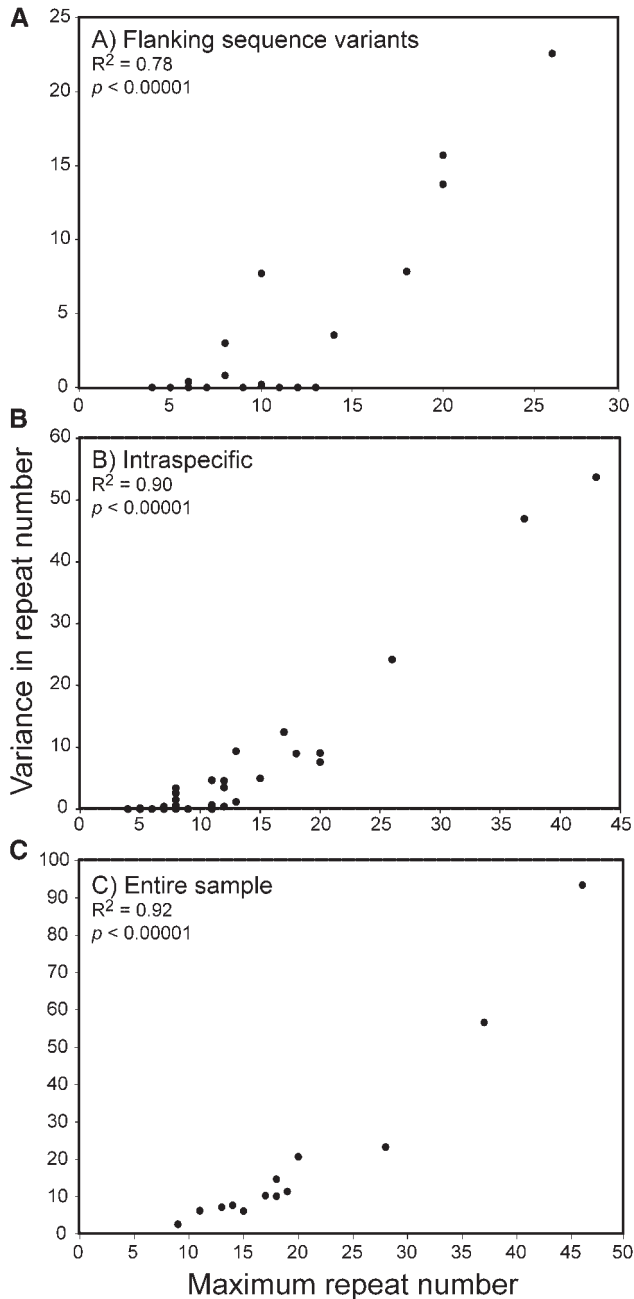
FIGURE 4.—Variance in repeat number plotted over maximum repeat number for comparisons at three genealogical depths: (A) Flanking-sequence variants ($n = 32$ variants). (B) Intraspecific ($n = 37$ species-by-locus combinations). (C) Entire sample ($n = 13$ loci). In all cases, the two variables are significantly correlated.

overlapping distributions at TML-ATC. At both QMA-TGG and QMA-CTTTT, NiA was fixed for a single repeat allele, whereas NcA was much more variable. To determine whether these distributions were more consistent with evolution under the SMM or TPM, we applied the likelihood method of NIELSEN (1997) to the five loci in each species. The SMM fit the data better than the TPM for 8 of the 10 locus-by-species combinations (Table 6). For the two comparisons in which the TPM was a better

fit, the proportion of multistep mutations was substantial (0.11) only for the DMG-CA locus in NcA. However, the gap between the smaller and larger alleles in the DMG-CA allele distributions in NcA (Figure 6) may not have been caused by multistep mutations. Rather, this gap may reflect the presence of two genealogical lineages of DMG-CA within NcA (see Figure 1), with different allele distributions for each genealogical lineage.

Mutation rates: The population mutation parameter θ ($\theta = 2N\mu$) can be estimated from data from each microsatellite locus. Because our loci were sampled from the same set of individuals, effective population size (N) was constant across loci within species, allowing us to directly compare relative mutation rates (μ) of loci. The θ-per-microsatellite locus in NcA and NiA was estimated using a likelihood approach (NIELSEN 1997; Tables 6 and 7). QMA-CTTTT and QMA-TGG were omitted from NiA calculations because they were not variable.

Microsatellite θ values, and therefore mutation rates, differed by an order of magnitude across loci within the same species (Table 7). In both NcA and NiA, θ for DMG-CA > TMI-AGC > TML-ATC. On average, microsatellite θ's for NcA were 77% greater than those for NiA, but this difference could be due to other factors, such as differences in effective population sizes of the two species.

The flanking regions were sampled from the same set of individuals as the microsatellites were, allowing for the comparison of the relative mutation rates of the two different, yet physically linked, DNA types. The θ per base pair of flanking DNA (Table 7) was estimated from the numbers of segregating sites (S; WATTERSON 1975). The ratio of microsatellite to flanking-sequence θ was quite variable among loci and between NcA and NiA. Interestingly, microsatellite θ's were on average $\sim$2500 times greater than flanking-sequence θ's for both NcA and NiA (2514 and 2498 times, respectively). Therefore, slippage rates of the microsatellites were $\sim$2500 times greater than the substitution rates of the nonrepetitive flanking DNA.

DISCUSSION

An advantage to a sequence-based study of microsatellite evolution is that the historical relationships among microsatellite alleles can be determined independently of their repeat number by the phylogenetic analysis of molecular variation observed in the flanking regions. The microsatellites and their associated flanking regions share the same evolutionary history and lines of descent due to their physical linkage. Much is known about the evolution of regular nonrepetitive DNA, allowing confident reconstructions of the genealogical relationships among the different variants of flanking sequence. Without flanking-sequence information, the phylogenetic affinities of microsatellite alleles can be deduced

```
Adjacent duplicated imperfections(locus,strain,species):
DMG-CA,D45,N.intermediaA:     CACACACACACACACACACACACACACACACACATATA
TML-ATC,D86,Phylo.Sp.2:       ATCATCATCATCATCATCAACAACAACAACAAC
QMA-TGG,D13,N.tetrasperma:    TGGTGGTGATGATGG
T1R-AGC,D124,N.sitophila:      AGCGGCGGCGGCAGCAGCAGCAGCAGCAGC
3L-AGC,D8,N.intermediaB:       AGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAGGAGG

Non-adjacent duplicated imperfections(locus,strain,species):
TML-ATC,D70,N.crassaB:        ATCAACATCATCATCATCATCATCAACATCATCATCATC
QMA-CTTTT,D111,N.crassaA:     CTTTTCTTTTCTTTTCTTTTCTTTTATCTCTTTTCTTTTATCTCTTTTCTTGTTGTTT
QMA-TGG,D1,N.intermedia:      TGGTAGTGGTGATGGTGATGG
T1R-AGC,D127,N.intermediaA:    AGCAGCAGCACCAGCACCAGCACCAACAGCAGCAGC
T7L-CGA,D19,N.crassaA:        CGACGACGACGGCGACGACGACGATGGCGACGACGACGGCGA
```

FIGURE 5.—Examples of adjacent and nonadjacent repeats with imperfections that have been duplicated within the same microsatellite allele. The duplicated repeat units with imperfections are in boldface type and underlined.

only from number of repeats, which may be a poor indicator of true relationships among long-diverged alleles. Placing the microsatellite mutations in the phylogenetic context of the more slowly evolving flanking regions reveals the complexities of the evolutionary dynamics of microsatellites. Intensive sequencing of the flanking regions provided novel approaches to the estimation of microsatellite mutational saturation and relative rates of mutation.

The sampling scheme of this study provided another novel approach: investigating microsatellite evolution at multiple genealogical depths and levels of divergence. By sampling multiple copies of the same flanking-sequence variant, the recent history of microsatellite evolution could be investigated. By sampling a moderate number of individuals from two species, *N. crassa* and *N. intermedia*, some species-level questions could be addressed. Including representatives from six other Neurospora species permitted comparisons of microsatellite evolution at a much deeper scale. The sampling scheme also allowed the conclusions drawn from the five intensively sampled microsatellite loci to be challenged by analysis of eight modestly sampled microsatellite loci.

Microsatellite variation: Nearly every study that has compared multiple microsatellite loci has reported differences in the levels of variability among loci (*e.g.*, Di RIENZO *et al.* 1994; ESTOUP *et al.* 1995a; GARZA *et al.* 1995; INNAN *et al.* 1997; HARR *et al.* 1998b; SCHUG *et al.* 1998). This study of microsatellites in Neurospora was no exception. The question is, Why are some loci more variable than others? The usefulness of microsatellite loci, regardless of their application, is dependent upon their variability. As such, it was this characteristic of microsatellites that we were most interested in predicting.

It is possible that microsatellites in genomic regions with higher mutation rates may show higher variability; however, the microsatellites and flanking regions examined here clearly evolve with different mutational dynamics. Microsatellite variability was not correlated with flanking region variability, demonstrating that microsatellite mutation rates are affected by different factors than are substitution rates in flanking regions. This study reports the first detailed investigation of the evolution of multiple, physically linked microsatellites. The fact that tightly linked microsatellites embedded within

the same genomic region could differ drastically in variability (*e.g.*, T7L-ACA and T7L-CGA) provided further evidence that genomic placement does not determine the variability of a particular microsatellite.

Microsatellite variability also differed among species, suggesting a possible species effect due to differences in effective population size or genome-wide mutation rates. However, no consistent species effects were detected, as exemplified by the fact that *N. intermedia* could be highly variable at one microsatellite, yet nearly monomorphic at another.

Of all the factors examined, the best predictors of microsatellite variability were mean and maximum repeat number. The association of increased variance with increased repeat number was quite robust and significant across all three genealogical depths (flanking-sequence variants, intraspecific, and entire sample; Figure 4). These findings are consistent with other studies that have found a significant positive relationship between variance and repeat number at single genealogical depths (GOLDSTEIN and CLARK 1995; INNAN *et al.* 1997; HUTTER *et al.* 1998; NEFF and GROSS 2001). Similar to results reported by GOLDSTEIN and CLARK (1995), variance in repeat number was best correlated with maximum repeat number. Several studies examining individual mutational events have also found that alleles with larger numbers of repeats are more likely to mutate than alleles with fewer numbers of repeats (*e.g.*, WEBER and WONG 1993; WIERDL *et al.* 1997; PRIMMER *et al.* 1998; SCHUG *et al.* 1998; ELLEGREN 2000a; VIGOUROUX *et al.* 2002; SHINDE *et al.* 2003).

This association is believed to exist for two reasons. During DNA replication, longer stretches of repeated units pose more of a problem to polymerase than do shorter stretches, making longer alleles more prone to slipped-strand mispairing. In addition, larger numbers of repeats provide more opportunities for misalignment during the reannealing of the nascent strand (see EISEN 1999). Extending the above arguments, there is expected to be a rough threshold of minimum repeat number below which a microsatellite is not likely to mutate or be variable. On the basis of the data presented here, microsatellites are expected to be variable within a Neurospora species if they have a maximum ≥ 7.70 and/or a mean ≥ 5.08 repeats. Other studies also have concluded that such a threshold exists (*e.g.*, MESSIER *et*
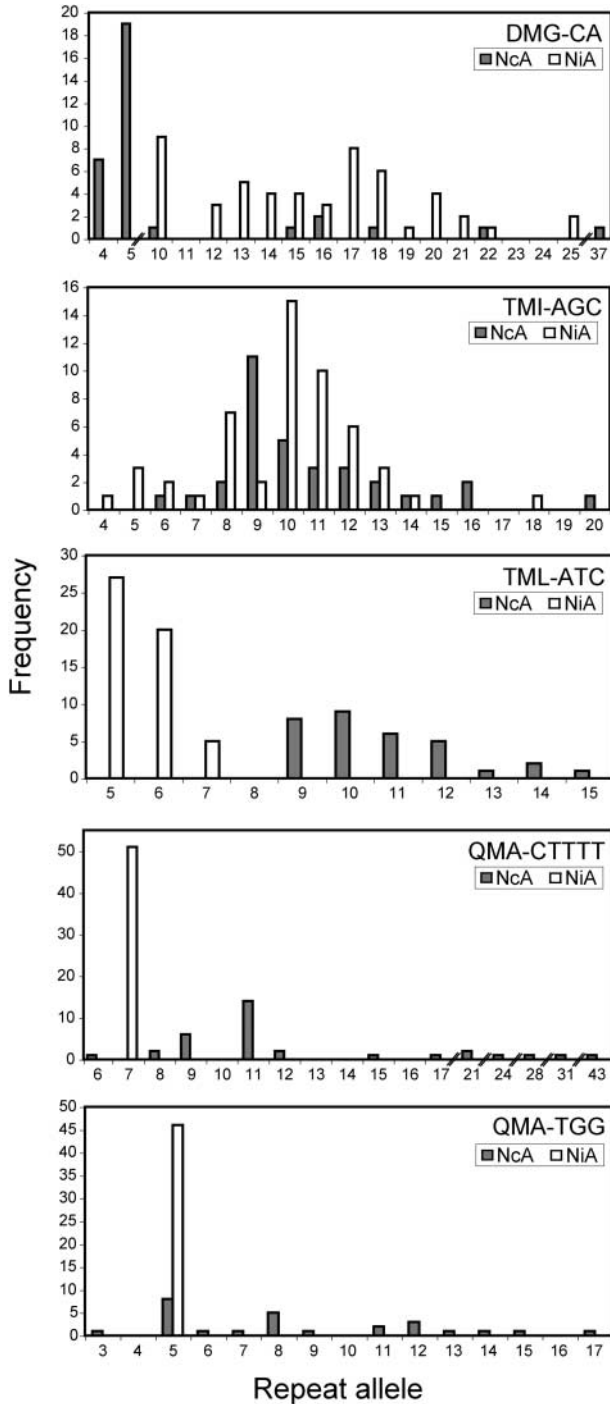
FIGURE 6.—Frequency distributions of repeat alleles at five microsatellite loci in *N. crassa* subgroup A (NcA) and *N. intermedia* subgroup A (NiA). Note that the *x*-axis is noncontinuous for some loci.

al. 1996; ROSE and FALUSH 1998; LAI and SUN 2003; SHINDE *et al.* 2003). For example, by examining slippage during PCR, SHINDE *et al.* (2003) estimated that the threshold was between four and eight repeats.

Phylogenetic utility of microsatellites: The ultimate consequence of extended periods of evolution via slipped-strand mispairing is deterioration of phyloge-

netic signal due to excessive homoplasy. Although the resolving power of microsatellites decreases with increasing divergence among groups, some studies have used these markers to estimate phylogenetic relationships among species (HARR *et al.* 1998a; PETREN *et al.* 1999; FISHER *et al.* 2000; ALVAREZ *et al.* 2001; IRION *et al.* 2003). These studies have had mixed results, in that microsatellite-based phylogenies do not necessarily agree with phylogenies based on other types of data.

Comparing the microsatellite-based trees with the previously reported interspecific phylogeny based on flanking sequences (DETTMAN *et al.* 2003a) allowed us to evaluate the phylogenetic utility of the microsatellites. The microsatellite loci we examined typically reached mutational saturation at flanking-sequence genetic distances that were less than genetic distances among species, and overlapping allele distributions among species were common. The high levels of interspecific homoplasy cast serious doubt on the efficacy of these microsatellites to reconstruct the phylogenetic relationships among Neurospora species, and trees constructed from the microsatellite data were not congruent with trees constructed from flanking-sequence data. These species have been evolving independently long enough for mutational saturation to have degraded the phylogenetic signal of microsatellites.

Given that microsatellite variability is dependent upon length, relatively short microsatellites may provide better resolution for the construction of interspecific phylogenies. However, it is unlikely that researchers could specifically choose microsatellites of the ideal length because loci are typically isolated from a single focal species, which may not be representative of the other species in the genus.

Mapping repeat alleles onto the genealogies of the flanking regions demonstrated that different flanking-sequence lineages within the same species can independently mutate to the same repeat allele, confirming microsatellite allelic homoplasy. Considering that microsatellites mutate at least three orders of magnitude faster than nonrepetitive DNA, two of the same repeat alleles that are embedded within different flanking-sequence variants are likely to be homoplastic. In fact, even identical repeat alleles associated with the same flanking-sequence variant may not be identical by descent because back mutations erase the evidence of previous mutations. When the sequences of the repeat arrays were considered, homoplasy was reduced because imperfect repeats can be used to distinguish among alleles with the same number of repeats. The substitution acts as a landmark for alignment and reveals if the same or different stretch of perfect repeats underwent slippage. The abundance of substitution mutations within the repeat arrays also demonstrated that microsatellite descriptors such as "perfect" and "imperfect" must be considered relative terms in reference to a particular group. Even though intraspecific variation is known to

TABLE 6

Results of likelihood-ratio tests comparing the fit of two mutational models to the data

Locus	Species	SMM θ	SMM Log $L(\theta)$	TPM θ	TPM P^a	TPM Log $L(\theta)$	Log L difference[b]	Best model
DMG-CA	NcA	105.47	−64.92	38.99	0.11	−40.69	24.23**	TPM
	NiA	46.94	−39.51	50.88	0.03	−42.06	2.55*	SMM
TMI-AGC	NcA	30.92	−25.11	35.77	0.03	−26.11	1.00	SMM
	NiA	25.36	−30.42	26.56	0.03	−28.37	2.05*	TPM
TML-ATC	NcA	14.51	−14.96	12.78	0.00	−15.14	0.18	SMM
	NiA	1.28	−8.40	1.14	0.00	−8.40	0.00	SMM
QMA-CTTTT	NcA	112.44	−79.02	117.12	0.10	−80.68	1.66	SMM
	NiA	0.00	NA	0.00	0.00	NA	NA	SMM
QMA-TGG	NcA	23.87	−28.53	41.41	0.05	−27.71	0.82	SMM
	NiA	0.00	NA	0.00	0.00	NA	NA	SMM

*$P < 0.05$; **$P < 0.005$.

[a] Estimated proportion of multistep mutations.

[b] One model fit the data significantly better than the other if twice the difference in log likelihood values was >3.84 (χ^2, d.f. = 1).

be high, species-wide conclusions are commonly drawn from the sequence information from a single individual. This study underscores the importance of sampling multiple individuals from the same species.

Intraspecific analyses: Microsatellites have been shown to be particularly useful for the investigation of population genetics (BRUFORD and WAYNE 1993; GOLDSTEIN and SCHLÖTTERER 1999) and continue to increase in popularity among most biological subdisciplines, including mycology (*e.g.*, BUCHELI *et al.* 2001; FISHER *et al.* 2001; ZHOU *et al.* 2001; BERGEMANN and MILLER 2002; DUNHAM *et al.* 2003). With this report, the first to characterize microsatellite loci in Neurospora, microsatellites can now be typed from large population samples (TURNER *et al.* 2001; JACOBSON *et al.* 2004) and used to investigate the population biology and ecology of this model filamentous fungus. We wanted to verify that these loci mutate in a manner consistent with the presently accepted models before initiating a large-scale

population genetic study. Within species, the data presented here are consistent with microsatellite evolution under the SMM with most mutational events involving changes of a single repeat unit. For two species examined, the estimated proportions of multistep mutations were large enough to reject a strict SMM for only one of the five loci investigated in detail. In contrast to interspecific analyses, no evidence for significant microsatellite mutational saturation within species was found, suggesting that these microsatellites will be useful for population genetic analyses. Although some intraspecific allelic size homoplasy was detected, it is not predicted to interfere with genetic distance or gene flow calculations. The efficacy of a microsatellite genetic distance to elucidate the true phylogenetic relationships among populations is dependent upon realistic assumptions of the mutation model. Because evidence suggests that intraspecific allelic diversity has been generated mainly under the SMM, SMM-based genetic distances

TABLE 7

θ estimates for five microsatellites and flanking regions

Species	Locus	Best microsatellite θ	Flanking regions bp	Flanking regions No. of sites	Flanking regions θ/bp	Microsatellite θ/flanking-sequence θ-ratio
NcA	DMG-CA	38.99	379.7	12	0.00779	5005.13
	TMI-AGC	30.92	445.9	23	0.01271	2432.73
	TML-ATC	14.51	538.3	32	0.01476	983.06
	QMA-CTTTT	112.44	463.4	62	0.03296	3411.41
	QMA-TGG	23.87	486.4	60	0.03232	738.55
	Five-locus mean	44.15	462.74	37.80	0.02011	2514.18
NiA	DMG-CA	46.94	373.1	16	0.00949	4946.26
	TMI-AGC	26.56	441.2	22	0.01104	2405.80
	TML-ATC	1.28	537.6	22	0.00906	141.28
	Three-locus mean	24.93	450.63	20.00	0.00986	2497.78

such as $(\delta\mu)^2$ are appropriate for use in future population genetic studies of Neurospora.

Mutation rates: The key feature of microsatellites as molecular markers is their hypermutability and, hence, their hypervariability in species and populations. Microsatellite mutation rates are estimated at 10^{-2}–10^{-6}/generation (ELLEGREN 2000b), which are several orders of magnitude greater than that of regular nonrepetitive DNA ($\sim 10^{-9}$, LI 1997). We found that microsatellite mutation rates were an average of $\sim$2500 times greater than the mutation rates of the flanking DNA. The best estimates of mutation rates in filamentous Ascomycetes were performed by KASUGA *et al.* (2002). They estimated that the mutation rate of noncoding DNA (introns) is between 1.12×10^{-9} and 1.00×10^{-8} substitutions/site/year, consistent with general estimates from other kingdoms. By simply multiplying the values of KASUGA *et al.* (2002) by 2500, the average mutation rate of these microsatellites is estimated to be between 2.80×10^{-6} and 2.50×10^{-5}. Although these are rough estimates, they are very similar to the microsatellite mutation rates estimated for *Saccharomyces cerevisiae* (1.2×10^{-5}, KRUGLYAK *et al.* 1998) and *Drosophila melanogaster* (2.8×10^{-6}, KRUGLYAK *et al.* 1998; 6.3×10^{-6}, SCHLÖTTERER *et al.* 1998; 9.3×10^{-6}, SCHUG *et al.* 1998).

Sequence data and assessment of microsatellite polymorphism: Microsatellite polymorphism in populations typically is assessed by electromorph length alone, but substitution mutations within the repeat array cannot be detected from electromorph length. Without sequencing these microsatellites in Neurospora, 33–56% of the total allelic variation would have been missed. A significant amount of variation in the nonrepetitive regions flanking the microsatellites was observed, even within a single species of Neurospora. Variation in flanking regions within species was commonly found in other systems (*e.g.*, JIN *et al.* 1996; GRIMALDI and CROUAU-ROY 1997; MACAUBAS *et al.* 1997; ORTÍ *et al.* 1997; COLSON and GOLDSTEIN 1999; BLANKENSHIP *et al.* 2002), but not all (ESTOUP *et al.* 1995b; ANGERS and BERNATCHEZ 1997; VIARD *et al.* 1998; CULVER *et al.* 2001). These substitutions in the flanking regions can reduce the efficiency of standard techniques for assessing microsatellite variation. For example, mutations in primer sites can cause amplification failure and null alleles. Even more troublesome are insertions and deletions because they can cause differences in electromorph length that are not attributable to differences in repeat number at the microsatellite. For intraspecific comparisons within NcA and NiA, the DMG, TMI, TML, and QMA flanking sequences contained an average of 1.75 multibase insertion/deletions that could obscure the detection of differences in repeat number between alleles (assuming the targeted electromorph length is up to 300 bp). If comparisons were made across species, even more multibase insertion/deletions would cause misinterpretation of electromorph length. Therefore, predicting

repeat number from electromorph length alone can be problematic due to undetected homoplasy. While there are clear benefits to determining repeat number directly from sequence data, doing so requires an unreasonably large amount of time and effort, even for population genetic studies of moderate size. A practical approach would be to sequence a small number of representative individuals from the groups under study to roughly assess levels of variation in the flanking regions and likelihood of allelic size homoplasy caused by insertion/deletions. This approach would allow for the exclusion of microsatellite loci that violate model assumptions or have complex mutational histories. Overall, it takes only a modest amount of sequencing to elucidate the general patterns of mutation and evolution at microsatellite loci. Before starting a large-scale population study, we recommend performing preliminary analyses to determine if the microsatellite loci are appropriate for addressing the particular questions of interest.

LITERATURE CITED

ALVAREZ, A. E., C. C. M. VAN DE WIEL, M. J. M. SMULDERS and B. VOSMAN, 2001 Use of microsatellites to evaluate genetic diversity and species relationships in the genus Lycopersicon. Theor. Appl. Genet. **103:** 1283–1292.

ANGERS, B., and L. BERNATCHEZ, 1997 Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol. Biol. Evol. **14:** 230–238.

BERGEMANN, S. E., and S. L. MILLER, 2002 Size, distribution, and persistence of genets in local populations of the late-stage ectomycorrhizal basidiomycete, *Russula brevipes*. New Phytol. **156:** 313–320.

BLANKENSHIP, S. M., B. MAY and D. HEDGECOCK, 2002 Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. Mol. Biol. Evol. **19:** 1943–1951.

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

BROHEDE, J., and H. ELLEGREN, 1999 Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. Proc. R. Soc. Lond. Ser. B **266:** 825–833.

BRUFORD, M. W., and R. K. WAYNE, 1993 Microsatellites and their applications to population genetic studies. Curr. Opin. Genet. Dev. **3:** 939–943.

BUCHELI, E., B. GAUTSCHI and J. A. SHYKOFF, 2001 Differences in population structure of the anther smut fungus *Microbotryum violaceum* on two closely related host species, *Silene latifolia* and *S. dioica*. Mol. Ecol. **10:** 285–294.

COLSON, I., and D. B. GOLDSTEIN, 1999 Evidence for complex mutations at microsatellite loci in Drosophila. Genetics **152:** 617–627.

CULVER, M., M. A. MENOTTI-RAYMOND and S. J. O'BRIEN, 2001 Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*). Mol. Biol. Evol. **18:** 1151–1156.

DETTMAN, J. R., D. J. JACOBSON and J. W. TAYLOR, 2003a A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. Evolution **57:** 2703–2720.

DETTMAN, J. R., D. J. JACOBSON, E. TURNER, A. PRINGLE and J. W. TAYLOR, 2003b Reproductive isolation and phylogenetic diver-

gence in *Neurospora*: comparing methods of species recognition in a model eukaryote. Evolution **57:** 2721–2741.

Dieringer, D., and C. Schlötterer, 2003 Microsatellite Analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes **3:** 167–169.

Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin *et al.*, 1994 Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA **91:** 3166–3170.

Dunham, S. M., A. Kretzer and M. E. Pfrender, 2003 Characterization of Pacific golden chanterelle (*Cantharellus formosus*) genet size using co-dominant microsatellite markers. Mol. Ecol. **12:** 1607–1618.

Eisen, J. A., 1999 Mechanistic basis for microsatellite instability, pp. 34–48 in *Microsatellites: Evolution and Applications*, edited by D. B. Goldstein and C. Schlötterer. Oxford University Press, Oxford.

Ellegren, H., 2000a Heterogeneous mutation processes in human microsatellite DNA sequences. Nat. Genet. **24:** 400–402.

Ellegren, H., 2000b Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. **16:** 551–558.

Estoup, A., and J. M. Cornuet, 1999 Microsatellite evolution: inferences from population data, pp. 49–65 in *Microsatellites: Evolution and Applications*, edited by D. B. Goldstein and C. Schlötterer. Oxford University Press, Oxford.

Estoup, A., L. Garnery, M. Solignac and J. M. Cornuet, 1995a 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., C. Tailliez, J. M. Cornuet and M. Solignac, 1995b Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol. Biol. Evol. **12:** 1074–1084.

Fisher, M. C., G. Koenig, T. J. White and J. W. Taylor, 2000 A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis*. Mol. Biol. Evol. **17:** 1164–1174.

Fisher, M. C., G. Koenig, T. J. White, G. San-Blas, R. Negroni *et al.*, 2001 Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration. Proc. Natl. Acad. Sci. USA **98:** 4558–4562.

FitzSimmons, N. N., C. Moritz and S. S. Moore, 1995 Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. Mol. Biol. Evol. **12:** 432–440.

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.

Gilbert, L., T. Kasuga, J. Townsend, N. L. Glass and J. W. Taylor, 2003 Comparative genomics within the genus *Neurospora*. Inoculum **54:** 22.

Goldstein, D., and A. Clark, 1995 Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucleic Acids Res. **23:** 3882–3886.

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

Goldstein, D. B., A. Ruiz-Linares, L. L. Cavalli-Sforza and M. W. Feldman, 1995a Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92:** 6723–6727.

Goldstein, D. B., A. Ruiz-Linares, L. L. Cavalli-Sforza and M. W. Feldman, 1995b An evaluation of genetic distances for use with microsatellite loci. Genetics **139:** 463–471.

Grimaldi, M.-C., and B. Crouau-Roy, 1997 Microsatellite allelic homoplasy due to variable flanking sequences. J. Mol. Evol. **44:** 336–340.

Harr, B., S. Weiss, J. R. David, G. Brem and C. Schlötterer, 1998a A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. Curr. Biol. **8:** 1183–1186.

Harr, B., B. Zangerl, G. Brem and C. Schlötterer, 1998b Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. Mol. Biol. Evol. **15:** 176–184.

Harr, B., B. Zangerl and C. Schlötterer, 2000 Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. Mol. Biol. Evol. **17:** 1001–1009.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hutter, C. M., M. D. Schug and C. F. Aquadro, 1998 Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol. Biol. Evol. **15:** 1620–1636.

Innan, H., R. Terauchi and N. T. Miyashita, 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics **146:** 1441–1452.

Irion, D. N., A. L. Schaffer, T. R. Famula, M. L. Eggleston, S. S. Hughes *et al.*, 2003 Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. J. Hered. **94:** 81–87.

Jacobson, D. J., A. J. Powell, J. R. Dettman, G. S. Saenz, M. M. Barton *et al.*, 2004 *Neurospora* in temperate forests of western North America. Mycologia **96:** 66–74.

Jin, L., C. Macaubas, J. Hallmayer, A. Kimura and E. Mignot, 1996 Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. Proc. Natl. Acad. Sci. USA **93:** 15285–15288.

Kasuga, T., T. J. White and J. W. Taylor, 2002 Estimation of nucleotide substitution rates in Eurotiomycete fungi. Mol. Biol. Evol. **19:** 2318–2324.

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

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

Kruglyak, S., R. T. Durrett, M. D. Schug and C. F. Aquadro, 1998 Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc. Natl. Acad. Sci. USA **95:** 10774–10778.

Lai, Y., and F. Sun, 2003 The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol. Biol. Evol. **20:** 2123–2131.

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

Li, W.-H., 1997 *Molecular Evolution*, pp. 177–213. Sinauer Associates, Sunderland, MA.

Macaubas, C., L. Jin, J. Hallmayer, A. Kimura and E. Mignot, 1997 The complex mutation pattern of a microsatellite. Genome Res. **7:** 635–641.

Makova, K. D., A. Nekrutenko and R. J. Baker, 2000 Evolution of microsatellite alleles in four species of mice (genus *Apodemus*). J. Mol. Evol. **51:** 166–172.

Messier, M., S. H. Li and C. B. Stewart, 1996 The birth of microsatellites. Nature **381:** 483.

Natvig, D. O., and G. May, 1996 Fungal evolution and speciation. J. Genet. **75:** 441–452.

Neff, B. D., and M. R. Gross, 2001 Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. Evolution **55:** 1717–1733.

Nielsen, R., 1997 A likelihood approach to populations samples of microsatellite alleles. Genetics **146:** 711–716.

Nielsen, R., and P. J. Palsbøll, 1999 Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. Mol. Phylogenet. Evol. **11:** 477–484.

Ortí, G., D. E. Pearse and J. C. Avise, 1997 Phylogenetic assessment of length variation at a microsatellite locus. Proc. Natl. Acad. Sci. USA **94:** 10745–10749.

Petren, K., B. R. Grant and P. R. Grant, 1999 A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proc. R. Soc. Lond. Ser. B **266:** 321–329.

Posada, D., and K. A. Crandall, 1998 ModelTest: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Primmer, C. R., and H. Ellegren, 1998 Patterns of molecular evolution in avian microsatellites. Mol. Biol. Evol. **15:** 997–1008.

Primmer, C. R., N. Saino, A. P. Møller and H. Ellegren, 1998 Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. Mol. Biol. Evol. **15:** 1047–1054.

Rose, O., and D. Falush, 1998 A threshold size for microsatellite expansion. Mol. Biol. Evol. **15:** 613–615.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SAS Institute, 1999 *JMPin*, Version 3.2.6. SAS Institute, Cary, NC.

SCHLÖTTERER, C., 2000 Evolutionary dynamics of microsatellite DNA. Chromosoma **109:** 365–371.

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

SCHLÖTTERER, C., B. AMOS and D. TAUTZ, 1991 Conservation of polymorphic simple sequence loci in cetacean species. Nature **354:** 63–65.

SCHLÖTTERER, C., R. RITTER, B. HARR and G. BREM, 1998 High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Mol. Biol. Evol. **15:** 1269–1274.

SCHUG, M., C. HUTTER, K. WETTERSTRAND, M. GAUDETTE, T. MACKAY *et al.*, 1998 The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Mol. Biol. Evol. **15:** 1751–1760.

SHIMODAIRA, H., and M. HASEGAWA, 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16:** 1114–1116.

SHINDE, D., Y. L. LAI, F. Z. SUN and N. ARNHEIM, 2003 Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: $(CA/GT)_n$ and $(A/T)_n$ microsatellites. Nucleic Acids Res. **31:** 974–980.

SHRIVER, M. D., L. JIN, R. CHAKRABORTY and E. BOERWINKLE, 1993 VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics **134:** 983–993.

SKUPSKI, M. P., D. A. JACKSON and D. O. NATVIG, 1997 Phylogenetic analysis of heterothallic *Neurospora* species. Fungal Genet. Biol. **21:** 153–162.

STRAND, M., T. A. PROLLA, R. M. LISKAY and T. D. PETES, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365:** 274–276.

SWOFFORD, D. L., 2001 *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4b10. Sinauer, Sunderland, MA.

TAKEZAKI, N., and M. NEI, 1996 Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics **144:** 389–399.

TURNER, B. C., D. D. PERKINS and A. FAIRFIELD, 2001 *Neurospora* from natural populations: a global study. Fungal Genet. Biol. **32:** 67–92.

TURNER, E., and J. W. TAYLOR, 2003 Marker development for a study of the genetic basis of reproductive isolation in *Neurospora*. Inoculum **54:** 49.

VALDES, A. M., M. SLATKIN and N. B. FREIMER, 1993 Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics **133:** 737–749.

VIARD, F., P. FRANCK, M. P. DUBOIS, A. ESTOUP and P. JARNE, 1998 Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. J. Mol. Evol. **47:** 42–51.

VIGOUROUX, Y., J. S. JAQUETH, Y. MATSUOKA, O. S. SMITH, W. F. BEAVIS *et al.*, 2002 Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. **19:** 1251–1260.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–275.

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

WIERDL, M., M. DOMINSKA and T. D. PETES, 1997 Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics **146:** 769–779.

ZARDOYA, R., D. M. VOLLMER, C. CRADDOCK, J. T. STREELMAN, S. KARL *et al.*, 1996 Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). Proc. R. Soc. Lond. Ser. B **263:** 1589–1598.

ZHOU, Z., M. MIWA and T. HOGETSU, 2001 Polymorphism of simple sequence repeats reveals gene flow within and between ectomycorrhizal *Suillus grevillei* populations. New Phytol. **149:** 339–348.

ZHU, Y., D. C. QUELLER and J. E. STRASSMANN, 2000 A phylogenetic perspective on sequence evolution in microsatellite loci. J. Mol. Evol. **50:** 324–338.

Communicating editor: D. BEGUN