

# A Test for Concordance Between the Multilocus Genealogies of Genes and Microsatellites in the Pathogenic Fungus *Coccidioides immitis*

M. C. Fisher,\* G. Koenig,† T. J. White,† and J. W. Taylor\*

\*Department of Plant and Microbial Biology, University of California at Berkeley; and †Roche Molecular Systems, Alameda, California

Uncovering the correct phylogeny of closely related species requires analysis of multiple gene genealogies or, alternatively, genealogies inferred from the multiple alleles found at highly polymorphic loci, such as microsatellites. However, a concern in using microsatellites is that constraints on allele sizes may occur, resulting in homoplasious distributions of alleles, leading to incorrect phylogenies. Seven microsatellites from the pathogenic fungus *Coccidioides immitis* were sequenced for 20 clinical isolates chosen to represent the known genetic diversity of the pathogen. An organismal phylogeny for *C. immitis* was inferred from microsatellite-flanking sequence polymorphisms and other restriction fragment length polymorphism-containing loci. Two microsatellite genetic distances were then used to determine phylogenies for *C. immitis*, and the trees found by these three methods were compared. Congruence between the organismal and microsatellite phylogenies occurred when microsatellite distances were based on simple allele frequency data. However, complex mutation events at some loci made distances based on stepwise mutation models unreliable. Estimates of times of divergence for the two species of *C. immitis* based on microsatellites were significantly lower than those calculated from flanking sequence, most likely due to constraints on microsatellite allele sizes. Flanking-sequence insertions/deletions significantly decreased the accuracy of genealogical information inferred from microsatellite loci and caused interspecific length homoplasies at one of the seven loci. Our analysis shows that microsatellites are useful phylogenetic markers, although care should be taken to choose loci with appropriate flanking sequences when they are intended for use in evolutionary studies.

## Introduction

Comparisons between multiple gene genealogies are increasingly used to detect reproductive isolation between taxa. Reduced gene flow due to geographic or reproductive isolation will leave its imprint in individual gene genealogies as drift sorts genetic variation. Subsequent comparisons of multilocus sequence data sets then are able to reveal the point at which reticulation between genealogies no longer occurs, thus diagnosing phylogenetic species (Templeton 1989; Avise and Ball 1990).

However, acquiring large multilocus sequence data sets with sufficient genetic variation to detect phylogenetic species can be time consuming, making the use of hypervariable markers attractive. Microsatellites have demonstrated their usefulness as genetic markers in studies of intraspecific population differentiation, but uncertainty exists as to their utility for describing phylogenetic species (Nauta and Weissing 1996; Ortí, Pearse, and Avise 1997; Paetkau et al. 1997). This uncertainty arises because constraints can occur on the range of allele sizes at microsatellite loci, limiting the genetic distance that can accrue between genetically isolated taxa (Garza, Slatkin, and Freimer 1995; Lehmann, Hawley, and Collins 1996; but see Kruglyak et al. 1998). A complicating factor is that the typically high mutation rates seen at microsatellite loci ( $\approx 10^{-2}$ – $10^{-5}$ , compared with  $\approx 10^{-9}$  for nucleotide substitutions; Dal-

las 1992; Weber and Wong 1993), coupled with size constraints in allelic distributions, will tend to counteract the effects of genetic drift with the reappearance of alleles that were previously lost from populations. This reappearance will result in homoplasy caused by alleles that are identical in size but not by descent, an effect that is expected to be strongest within populations of large effective size, such as those often found for species of microbes. For microsatellites to be useful as genetic markers in studies of morphologically depauperate species, they need to be able to diagnose cryptic species, as well as illuminate intraspecific relationships (Taylor et al. 1999). However, it is uncertain to what extent they meet these conditions.

A common approach for determining whether microsatellites are able to reconstruct species level relationships has been to use simulations of the evolution of microsatellite genetic distances over time. Microsatellites differ from classical genetic markers in that the origin of novel alleles is best described by the stepwise mutation model (SMM; Ohta and Kimura 1973) rather than the infinite-alleles model, because the principal mode of evolution is the accumulation of length polymorphisms rather than substitutions (Levinson and Gutman 1987a, 1987b). Several genetic distances have been devised that specifically take into account the distribution of allele size as well as frequency, for instance,  $D_{SW}$  (Shriver et al. 1995),  $R_{ST}$  (Slatkin 1995),  $D_1$ , and the related distance  $(\delta\mu)^2$  (Goldstein et al. 1995a, 1995b), while other distances, such as  $D_{AS}$  (Stephens et al. 1992), are based simply on allele frequency. Simulations have shown that genetic distances based on allele frequencies are generally more accurate at reconstructing relationships for low to moderate levels of genetic divergence, but these statistics rapidly approach an asymptote and lose resolution at larger distances (Gold-

Abbreviations: CA, California; non-CA, non-California; PHT, partition homogeneity test; SMM, stepwise mutation model.

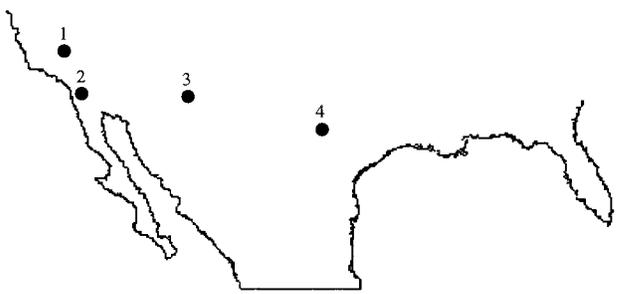
Key words: microsatellite, *Coccidioides immitis*, gene genealogy, molecular evolution, multilocus, genotype.

Address for correspondence and reprints: M. C. Fisher, Department of Plant and Microbial Biology, University of California at Berkeley, Berkeley, California 94720-3102. E-mail: matthewfisher@email.com.

*Mol. Biol. Evol.* 17(8):1164–1174. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

**Table 1**  
***Coccidioides immitis* Isolates Used in this Study and the Approximate Geographical Areas Where They Were Isolated**



GEOGRAPHICAL AREA			
1. Bakersfield, Calif.	2. San Diego, Calif.	3. Tucson, Ariz.	4. San Antonio, Tex.
2002 BA (CA)	2102 SD (CA)	1036 AZ (non-CA)	2127 TX (non-CA)
2009 BA (CA)	2105 SD (CA)	1038 AZ (non-CA)	2128 TX (non-CA)
2012 BA (CA)	2395 SD (CA)	1039 AZ (non-CA)	2129 TX (non-CA)
2267 BA (CA)		1040 AZ (non-CA)	2133 TX (non-CA)
2275 BA (CA)		1042 AZ (non-CA)	
2281 BA (CA)		1052 AZ (non-CA)	
		1443 AZ (non-CA)	

NOTE.—Roche Molecular Systems Culture Collection (RMSCC) reference numbers are followed by phylogenetic species designation based on RFLP analysis (in parentheses).

stein et al. 1995a). On the other hand, under conditions of unconstrained mutation, distances that utilize allele size as well as frequency tend to reconstruct deeper divergences better and will show a linear relationship between genetic distance and time (Goldstein et al. 1995a). A general conclusion of all studies is that unless large numbers of loci (over 50) are used, confidence in predicting the correct phylogeny is low (Garza, Slatkin, and Freimer 1995; Nauta and Weissing 1996; Takezaki and Nei 1996).

Given the uncertainty in the mode and tempo of the evolution of microsatellite genetic distances and a lack of empirical data from many taxa, such as microbes, it is important to determine the performance of these markers over large genetic distances. Here, we analyze a microsatellite data set collected from the human fungal pathogen *Coccidioides immitis*, the etiological agent of coccidioidomycosis (San Joaquin Valley fever). We use this organism because it exhibits a clear demonstration of cryptic speciation and phylogeographic isolation at several scales in the southwestern United States. The use of gene genealogies has shown that the fungus consists of two apparently allopatric cryptic species (Koufopanou, Burt, and Taylor 1997, 1998). These have been genetically isolated for an estimated 12 Myr, and within each species there is a strong pattern of geographical genetic isolation (Burt et al. 1997; M. C. Fisher, unpublished data). In this paper, we test the utility of microsatellite loci in recovering these relationships.

## Materials and Methods

### Population Samples

Seventeen clinical isolates of *C. immitis* were chosen that represented the known cryptic species and phy-

logeographically isolated populations from the southwestern United States based on data from previous studies (Burt et al. 1997; Koufopanou, Burt, and Taylor 1997). In addition, three clinical isolates of previously unknown genotype were included from San Diego, Calif. (2102SD, 2105SD, and 2395SD). The isolates, identification numbers, genotype, and the geographic areas from which they were isolated are shown in table 1. Each isolate was cultured in liquid media within a BL3 containment facility, was autoclaved to kill the mycelia, and had its DNA extracted according to Burt et al. (1995) for use as the template in a PCR.

### Microsatellite Loci

Seven dinucleotide microsatellites were used in this study; their isolation and accession numbers have been described elsewhere (Fisher et al. 1999). Four of these microsatellites (*621.2*, *ACJ*, *GAI*, and *GAC2*) were originally isolated from the California (CA) species of *C. immitis*, and three (*KO1*, *KO7*, and *KO9*) were isolated from the non-California (non-CA) species. Each locus was amplified by PCR from the 20 fungal isolates using the conditions and primers described by Fisher et al. (1999).

### Microsatellite Flanking-Sequence Genealogies

Microsatellite-containing loci were cycle-sequenced using fluorescently labeled dye terminators (Amersham), read using an automated sequencer (Applied Biosystems), and the sequences were aligned using the CLUSTAL W option in Sequence Navigator (Applied Biosystems). The microsatellite repeat motifs were

excluded, and parsimony analyses were performed on each of the flanking sequences using the PAUP\*, version 4.0b2a, software package (Swofford 1998). Indels were coded and treated as single characters, and parsimony analyses were performed using heuristic searches of the data set with the tree bisection-reconnection branch-swapping option on and the steepest-descent option off. Bootstrap consensus trees were constructed by performing heuristic searches on 1,000 bootstrap-resampled data sets. In addition, distance analyses were performed. Because the sequences used here are closely related, corrections for multiple hits proved to be unnecessary, and therefore an uncorrected distance was used. Pairwise distance estimates were grouped using the neighbor-joining algorithm in PAUP\*.

The coalescence time for the combined genealogies of all seven loci in numbers of generations,  $\tau_{\text{gen}}$ , was calculated from the microsatellite-flanking sequences using

$$\tau_{\text{gen}} = K/2r, \quad (1)$$

where  $K$  is the sequence divergence between populations minus the pairwise sequence divergence within populations, and  $r$  is the nucleotide substitution rate (estimated as  $1 \times 10^9$  per nucleotide per generation; Li and Graur 1991). Here, populations were defined as isolates coming from a single geographical region, either Bakersfield, Calif., San Diego, Calif., Tucson, Ariz., or San Antonio, Tex. (table 1).

#### Independent Restriction Fragment Length Polymorphism-Containing Loci

Due to the nonindependence between these microsatellite loci and their flanking sequences as a consequence of physical linkage, an independent phylogeny of these isolates of *C. immitis* was obtained by analyzing seven loci that contained polymorphic restriction sites. Loci were chosen which contained restriction fragment length polymorphisms (RFLPs) in either the CA species (loci *Vl*, *IT*, and *Ra*; Fisher et al. 2000) or the non-CA species (loci *z*, *bl*, *bq*, and *eI*; Burt et al. 1997). Each locus was amplified for all isolates, and the polymorphic site was detected by digestion with the appropriate restriction endonuclease as previously described (*Vl*, *HinfI*, *BstNI*; *IT*, *HaeIII*; *Ra*, *NruI* [Fisher et al. 2000]; *z*, *HinfI*; *bl*, *DdeI*; *bq*, *NsiI*; *eI*, *BsmI* [Burt et al. 1997]). Phylogenetic analyses of the RFLP data set were performed using PAUP\* as described above. Congruence between each flanking-sequence genealogy and the RFLP data set was tested using the partition homogeneity test (PHT; Farris et al. 1995; Huelsenbeck, Bull, and Cunningham 1996), implemented in PAUP\*. This test shuffles phylogenetically informative sites between data partitions (here, each flanking-sequence locus and the RFLP data set), finding the sum of the lengths of the most parsimonious trees for each replicate, then comparing these lengths with the summed tree lengths of the original data partitions. If the sum for the original data is significantly shorter, the null hypothesis of congruence can be rejected.

#### Measurements of Microsatellite Genetic Distance

We selected two commonly used measures of microsatellite genetic distance for analysis. The first,  $D_{AS}$ , is based on simple allele frequency data and calculates multilocus pairwise distance measurements as  $1 - (\text{the total number of shared alleles at all loci}/n)$ , where  $n$  is the number of loci compared (Stephens et al. 1992; Bowcock et al. 1994). The second,  $(\delta\mu)^2$ , was developed specifically for microsatellite applications and assumes a single-step SMM.  $(\delta\mu)^2$  is the square of the difference in mean allele size ( $x$ ) between two populations *A* and *B* such that  $(\delta\mu)^2 = (x_A - x_B)^2$  (Goldstein et al. 1995b). In populations where mutation-drift equilibrium can be assumed, this distance is linear with respect to time, as well as independent of population size (Goldstein et al. 1995b). For pairwise distances between individuals,  $(\delta\mu)^2$  was calculated as the average squared difference in allele size, a distance that is functionally identical to the measure D1 used by Goldstein et al. (1995b) and, where used, will be denoted as such.

Microsatellite distances were calculated for (1) the numbers of repeats within a locus and (2) the absolute length of the locus (therefore including the length variation introduced by flanking-sequence indels). Confidence intervals for  $D_{AS}$  and  $(\delta\mu)^2$  were calculated by bootstrapping over loci using the program MICROSAT (Minch et al. 1995), and neighbor-joining trees of both distances were then constructed in PHYLIP, version 3.5c (Felsenstein 1991).

#### Comparisons of Different Tree Topologies

Tree topologies were compared against one another using the Kishino-Hasegawa test (Kishino and Hasegawa 1989). This test compares the log likelihoods of each tree topology and uses a paired *t*-test to reject topologies that are significantly less likely than the optimal tree. Analyses were performed using the Hasegawa-Kishino-Yano model of mutation (Hasegawa, Kishino, and Yano 1985) with base frequencies and transition-transversion ratios empirically determined from the data sets.

In order to test the power of the gene sequence (flanking-sequence and RFLP data sets) to reject the hypothesis of congruency between the microsatellite and gene sequence trees, three sets of randomized trees were created. These sets of trees were then compared against the original gene sequence tree using the Kishino-Hasegawa test. The sets of trees were (1) non-CA isolates randomized within the non-CA clade, (2) CA isolates randomized within the CA clade, and (3) all isolates randomized within and between the CA and non-CA clades. For each set, 100 randomized trees were generated with the "generate trees" option in PAUP\*.

#### Time to Coalescence of Microsatellite Alleles

The coalescence time between populations inferred from microsatellite variation was calculated following Goldstein et al. (1995b), who demonstrated that the time to coalescence for microsatellite alleles is related to generation time by

A.												
LOCUS: <i>GAI</i>	<i>ACJ</i>			<i>GAC2</i>			<i>621</i>		<i>KOI</i>			
Character	11111111122222	22222	2	33	3333	3	33	3444444	4	4445 555	5	55
1	2345678901234567890123	45678	9	01	2345	6	78	9012345	6	7890 123	4	56
<b>ISOLATE</b>												
2002BA	(CT)13TA0TAA11T10TCCC1AG1AAA	A1CCC	(GT)17	0G	CG0T	(CT)18??	AGG010G	(CT)13	C0GT	ATG	(TG/AG)21	A0
2009BA	(CT)13		(GT)20			(CT)15??	...1...	(CT)12			(TG/AG)22	
2012BA	(CT)16	T	(GT)19			(CT)17??	...A	(CT)15			(TG/AG)22	
2267BA	(CT)13		(GT)18		?	(CT)16GT	...1...	(CT)18			(TG/AG)21	
2275BA	(CT)14	0	(GT)26			(CT)17GT	...A	(CT)16			(TG/AG)21	
2281BA	(CT)13		(GT)20			(CT)17GT	?1	(CT)17			(TG/AG)21	
2102SD	(CT)13		(GT)21			(CT)18GT	...0	(CT)14			(TG/AG)21	
2105SD	(CT)15	A	(GT)20			(CT)19GT	...1	(CT)14			(TG/AG)21	
2395SD	(CT)14		(GT)201			(CT)16GT	...0	(CT)14			(TG/AG)21	
1036AZ	(CT)11CT1CCT00G01ATT.0.AO	GGC	C0..T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)19C1
1038AZ	(CT)11CT1CCT00G01.TTA0.AO	GGC	C0..T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)18C1
1039AZ	(CT)11CT1CCT00G01ATT?0.AO	GGC	C0..T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)19C1
1040AZ	(CT)11CT1CCT00G01ATT.0.AO	GGC	C0..T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)19C1
1042AZ	(CT)12CC1CCT.0G01.TT.0GAO	GGC	C0.TT	(GT)111A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	AAA	(TG/AG)17C1
1052AZ	(CT)11CT1CCT.0G01.TTA0.AO	GGC	C0.TT	(GT)111A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)19C1
1443AZ	(CT)12CT1CCT00G01.TT.0GAO	GGC	C0.TT	(GT)111A	GC1C	(CT)8	AG	??????????	????	GAA	(TG/AG)19C1	
2127TX	(CT)12CT1CCT00G01ATT.0.AO	GGC	C.G.T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)20C1
2128TX	(CT)12CT1CCT00G01ATT.0.AO	GGC	C.G.T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)18C1
2129TX	(CT)12CT1CCT00G01ATT?0.AO	GGC	C.G.T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)19C1
2133TX	(CT)12CT1CCT00G01ATT.0.AO	GGC	C.G.T	(GT)101A	GC1C	(CT)8	AG	GAA1.1.	(CT)6	G1AC	GAA	(TG/AG)18C1

B.										
LOCUS: <i>KO3</i>	<i>KO7</i>			LOCUS: <i>z</i>	<i>bl</i>	<i>bq</i>	<i>e1</i>	<i>Vl</i>	<i>IT</i>	<i>Ra</i>
Character	5	5	5666666666667777	77	7	77				
7	8	901234567890123	45	6	78					
<b>ISOLATE</b>										
2002BA	C	(TC/GT)19	TGGTGCGTA0TCCA	0T	(CT)9	0A				
2009BA	.	(TC/GT)19			(CT)9					
2012BA	.	(TC/GT)19			(CT)8	1.				
2267BA	.	(TC/GT)19			(CT)10					
2275BA	.	(TC/GT)19			(CT)9					
2281BA	.	(TC/GT)19			(CT)9					
2102SD	.	(TC/GT)19			(CT)9					
2105SD	.	(TC/GT)19			(CT)9					
2395SD	.	(TC/GT)19			(CT)8	1.				
1036AZ	A	(TC/GT)21	0CCCGTTTGG1CTTG	1C	(CT)12	1G				
1038AZ	A	(TC/GT)26	0CCCGTTTGG1CTTG	1.	(CT)9	1G				
1039AZ	A	(TC/GT)20	0CCCGTTTGG1CTTG	1C	(CT)12	1G				
1040AZ	A	(TC/GT)29	0CCCGTTTGG1CTTG	1.	(CT)7	.G				
1042AZ	A	(TC/GT)21	0CCCGTTTGG1CTTG	1C	(CT)11	1G				
1052AZ	A	(TC/GT)21	0CCCGTTTGG1CTTG	1C	(CT)12	1G				
1443AZ	A	(TC/GT)21	0CCCGTTTGG1CTTG	1.	(CT)8	1G				
2127TX	A	(TC/GT)20	0CCCGTTTGG1CTTG	1.	(CT)7	.G				
2128TX	A	(TC/GT)22	0CCCGTTTGG1CTTG	1.	(CT)8	.G				
2129TX	A	(TC/GT)20	0CCCGTTTGG1CTTG	1.	(CT)8	.G				
2133TX	A	(TC/GT)22	0CCCGTTTGG1CTTG	1.	(CT)7	.G				

FIG. 1.—Character states for polymorphisms in (A) microsatellite-containing loci *GAI* (positions 1–23), *ACJ* (24–31), *GAC2* (32–38), *621* (39–50), *KOI* (51–56), *KO3* (57–73), and *KO7* (74–78) and (B) RFLP-containing loci *z*, *bl*, *bq*, *e1*, *Vl*, *IT*, and *Ra*. Characters in bold are microsatellite motifs followed by the numbers of repeats within the isolate. Indels are coded as “0/1,” and the presence/absence of a restriction site is coded as “+/-”. Alleles identical to those in the reference individual (2002BA) are scored as “.”.

$$(\delta\mu)^2 = 2\mu\tau_{gen} \tag{2}$$

The dinucleotide microsatellite mutation rate,  $\mu$ , was estimated as  $2.6 \times 10^{-5}$  from the values recently determined for yeast (Kruglyak et al. 1998).

**Results**  
Phylogenetic Analysis of Microsatellite-Flanking Sequence and RFLPs

DNA flanking the seven microsatellite loci was sequenced for the 20 isolates of *C. immitis*, resulting in 2,176 bp of nucleotide sequence (excluding the microsatellites themselves). Within this sequence, indels were

treated as single characters, resulting in a final data set of 2,147 characters. Of these, 93 sites were polymorphic, of which 78 were parsimony-informative. The seven RFLP-containing loci contained nine polymorphic sites (loci *z* and *vl* each contained two polymorphic sites), of which seven were parsimony-informative. Multilocus genotypes of all *C. immitis* isolates and the positions of the microsatellites within each locus are shown in figure 1.

Maximum-parsimony analyses of the sequences flanking the seven microsatellite-containing loci indicated that all but one, *KO7*, supported a strong central branch separating CA from non-CA isolates with >87%

**Table 2**  
**Maximum-Parsimony Analyses of Individual**  
**Microsatellite-Flanking Loci**

Locus	No. of Steps Separating CA and Non-CA	% Bootstrap Support Separating CA and Non-CA
<i>GAI</i> .....	31	100
<i>ACJ</i> .....	8	97
<i>GAC2</i> .....	6	99
<i>621</i> .....	17	87
<i>KO1</i> .....	6	92
<i>KO3</i> .....	19	100
<i>KO7</i> .....	5	<50

bootstrap support (table 2). The PHT test showed that the summed tree lengths of each data partition (each locus) were not significantly different from those from 1,000 shuffled data sets, demonstrating that conflict among the seven microsatellite-flanking sequences was not significant ( $P = 0.73$ ). Accordingly, all loci were combined into a single data set.

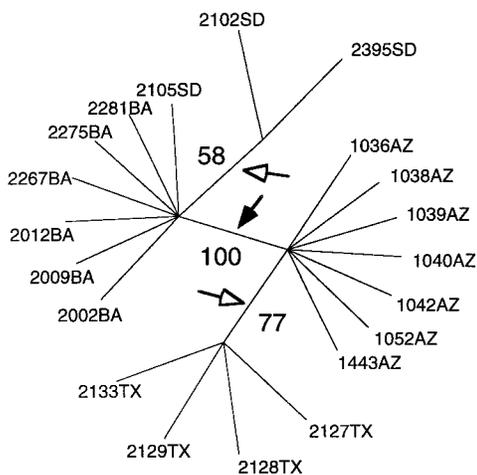
Heuristic searches of the flanking-sequence and RFLP data sets using parsimony found 2,437 trees of 101 steps and 162 trees of 13 steps, respectively. The strict consensus tree for each data set is shown in figure 2. Bootstrap analysis supported a strong central branch separating CA from the non-CA genotypes for both data sets, with the three San Diego isolates of unknown genotype being placed in the CA clade for both consensus trees. The RFLP data set showed no further resolution, with all isolates falling into two polytomies, the CA and the non-CA clades, supporting the previous designation of these clades as phylogenetic species (Koufopanou, Burt, and Taylor 1997, 1998; Fisher et al. 2000). The

combined microsatellite flanking-sequence data set showed more resolution, with Texas isolates forming a unique clade (supporting  $F_{ST}$  analysis of RFLP data; Burt et al. 1997) and two of the three San Diego isolates clustering with weak bootstrap support. Combinability of the flanking-sequence and RFLP data sets was tested using the PHT test. Summed tree lengths were no different from those from 1,000 shuffled data sets ( $P = 0.375$ ), showing that conflict between the two data sets was not significant, and they were therefore combined. This combined data set is referred to henceforth as the gene sequence data set. Maximum-parsimony analysis of the gene sequence data set found 462 trees of 117 steps and showed increased support for the San Antonio clade. However, San Diego isolates did not cluster with significant bootstrap support, suggesting that differentiation between the Bakersfield and San Diego populations is low.

Distance analysis of the gene sequence data set was used to infer a second tree topology. The Kishino-Hasagawa test showed that none of the 462 parsimony trees was significantly different from the tree built using the distance method. Therefore, to standardize the method of inferring tree topology with that used for the microsatellite data, the gene sequence distance tree was taken as the organismal phylogeny against which the microsatellite trees were tested.

As a means of finding the amount of support within this data set for the various clades, randomized trees were constructed as described in *Materials and Methods*. Trees randomizing isolates within the CA clade were not significantly different with respect to the original tree (table 3). However, randomizing isolates within the non-CA clade resulted in a large number of trees

**A. Flanking sequence consensus tree**  
 (characters 2-78)



**B. RFLP consensus tree**  
 (characters 79-87)

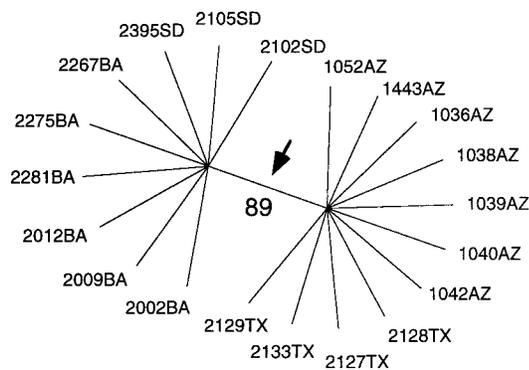


FIG. 2.—A, Strict consensus of the 2,437 unrooted most-parsimonious trees found from the sequence flanking the seven microsatellite loci. The dark arrow shows the strongly supported branch separating the two phylogenetic species CA and non-CA. The light arrows show phylogenetically isolated populations (San Diego and San Antonio) with weaker bootstrap support. B, Maximum-parsimony analysis of the same 20 isolates using biallelic restriction fragment length polymorphism (RFLP) data from seven independent loci.

**Table 3**  
**Kishino-Hasegawa Test Results**

Tree	−ln L	Difference in −ln L	T	P
Gene sequence distance tree . . . . .	3,458	Best tree	—	—
$D_{AS}$ tree (length of repeats only) . . . . .	3,487	29	1.17	0.24
$D_{AS}$ tree (length of entire locus) <sup>a</sup> . . . . .	3,496	39	1.79	0.07
D1 tree (length of repeats only) . . . . .	3,542	84	1.87	0.06
D1 tree (length of entire locus) <sup>a</sup> . . . . .	3,548	89	1.86	0.06
Non-CA, randomized <sup>b</sup> (SD) . . . . .	3,536 (16.07)	53	1.81 (0.33)	0.09 (0.09)
CA, randomized <sup>c</sup> (SD) . . . . .	3,482 (5.58)	23	1.16 (0.14)	0.25 (0.07)
Non-CA/CA, randomized <sup>d</sup> (SD) . . . . .	4,983 (266)	1,524	8.17 (0.06)	0.00 (0.00)

NOTE.—P is significant if <0.05. SD = standard deviation.

<sup>a</sup> Microsatellite distances were inferred from the absolute length of a locus rather than from the number of repeats at a locus.

<sup>b</sup> Means found when 100 trees, with non-CA isolates randomized within the non-CA clade, were compared against the gene sequence tree.

<sup>c</sup> Means found when 100 trees, with CA isolates randomized within the CA clade, were compared against the gene sequence tree.

<sup>d</sup> Means found when 100 trees, with all isolates randomized within and between the CA and non-CA clades, were compared against the gene sequence tree.

with significantly worse topologies, and randomizing between the CA and the non-CA clades resulted in all trees being worse. Overall, these Kishino-Hasegawa test results show that there is adequate phylogenetic structure within this data set to illuminate incongruencies between different tree topologies if the two main supported branches (Texas/non-CA and CA/non-CA) are affected.

#### Variation in Microsatellite-Containing Loci

Variation in the dinucleotide repeats of the seven microsatellite loci is shown in figure 3A. Six of the seven microsatellites show nonoverlapping species-specific allele distributions between CA and non-CA *C. immitis* (all but locus *KO7*; fig. 3A). Ascertainment bias occurs when microsatellites are longer in the species from which they were found than in conspecifics (Ellegren et al. 1997). For these loci in *C. immitis*, this bias is strong. All four microsatellites cloned from CA had more repeats compared with those found in non-CA *C. immitis*, and the converse comparison showed that two of the three loci were longer in the focal species, non-CA, than in CA.

Figure 3B shows the entire length of each locus and therefore includes length variation of the flanking sequence as well as the microsatellite itself. Comparison of figure 3A and B reveals that all seven loci show interspecific size variation in the microsatellite-flanking sequence, caused by indels. This size variation has led to homoplasy in locus *KO3*, where alleles are no longer species-specific (fig. 3B). Intraspecific homoplasy is also apparent at locus *KO7*, where alleles have converged in size within both the CA and the non-CA *C. immitis* due to the occurrence of flanking-sequence indels. Two loci (*GAI* and *621*) show intraspecific size polymorphisms in the flanking sequence. These loci are responsible for increasing the numbers of alleles found within the CA (loci *GAI* and *621*) and the non-CA (locus *GAI*; fig. 3B) *C. immitis* and are principally due to single-nucleotide variation at short mononucleotide (T)<sub>n</sub> and (G)<sub>n</sub> motifs. Microsatellite-flanking sequence is often rich in these arrays, and they are prone to indels, a phenomenon that has been reported by other workers (Ortí, Pearse, and Avise 1997; Colson and Goldstein 1999). This high level of flanking-sequence indel events directly dem-

onstrates the necessity of sequencing several microsatellite alleles to ascertain homology, as it is apparent here that flanking-sequence diversity is directly inflating measurements of microsatellite genetic diversity.

#### Reconstruction of the Organismal Phylogeny Using Microsatellite Distance Measurements $D_{AS}$ and D1

The distances  $D_{AS}$  and D1 were calculated from the numbers of repeats at each locus and were used with the neighbor-joining algorithm to cluster the *C. immitis* isolates (fig. 4B and C). Both distances fully resolved the two CA/non-CA *C. immitis* species with high bootstrap support, and  $D_{AS}$  grouped the San Antonio isolates as a single clade within the non-CA clade, albeit with less bootstrap support than was found in the gene sequence tree. However, use of D1 scattered the San Antonio isolates within the non-CA clade (fig. 4C). Comparison of the log likelihood value of the  $D_{AS}$  tree against that found from the gene sequence tree, using the Kishino-Hasegawa test, showed no significant difference in tree topologies (table 3). The D1 distance performed less well, and the tree topology inferred from it was much closer to being significantly less likely when compared with the gene sequence tree.

We tested the effect of flanking-sequence indels on the topology of the inferred microsatellite genealogies by using the length variation at each locus, rather than the number of repeats at a locus, as an allele. In this case, log likelihoods for both  $D_{AS}$  and D1 trees were worse than those inferred from the numbers of repeats (table 3). This result demonstrates that flanking-sequence length variation is responsible for a decrease in the accuracy of the microsatellite distances for recovering the correct genealogy.

#### Dating the Split of *C. immitis* Lineages

The pairwise divergence between CA and non-CA *C. immitis* for flanking nucleotide substitutions was measured as  $2.55 \times 10^{-2}$ , a value that compares reasonably closely with that previously estimated from third-position variation within *C. immitis* coding genes ( $1.6 \times 10^{-2}$ ; Koufopanou, Burt, and Taylor 1998). Based on our value, we estimate that divergence be-

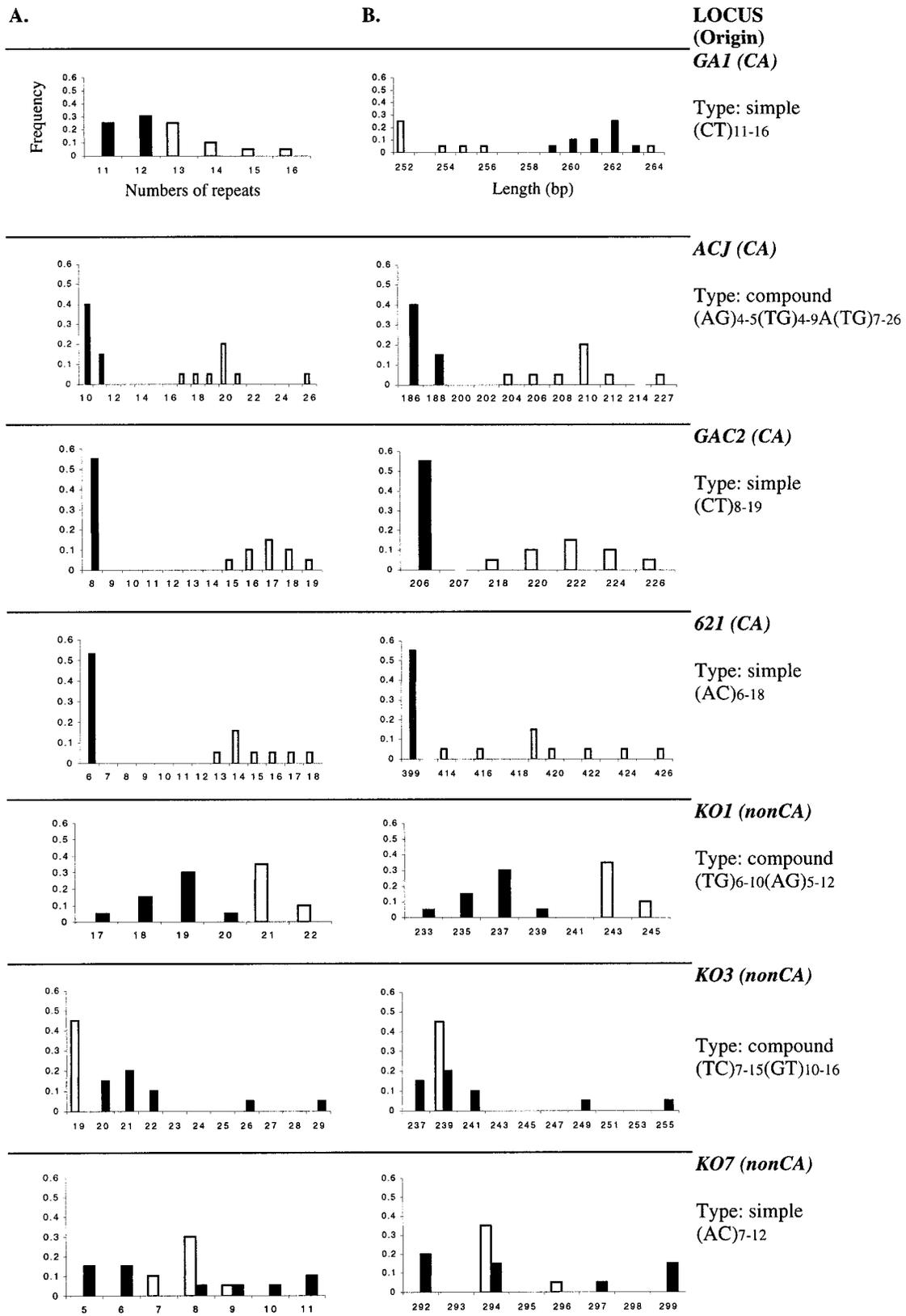
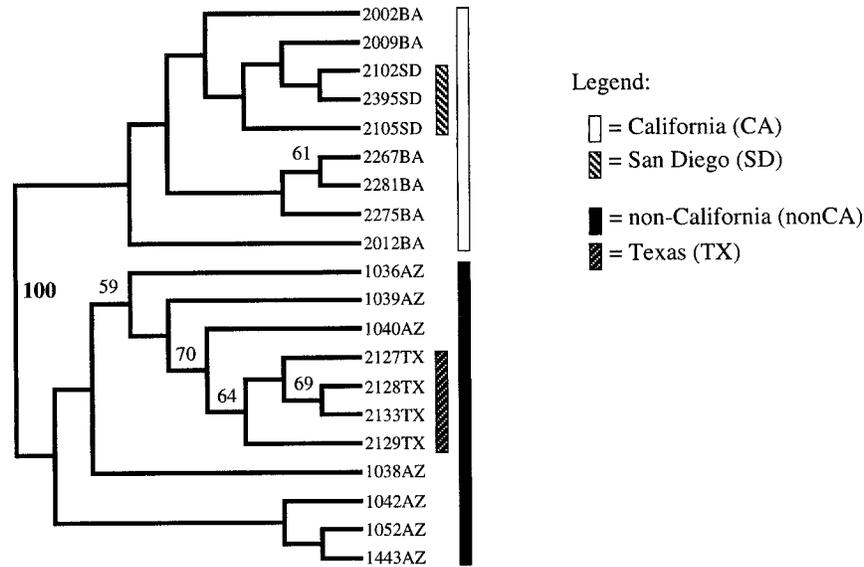
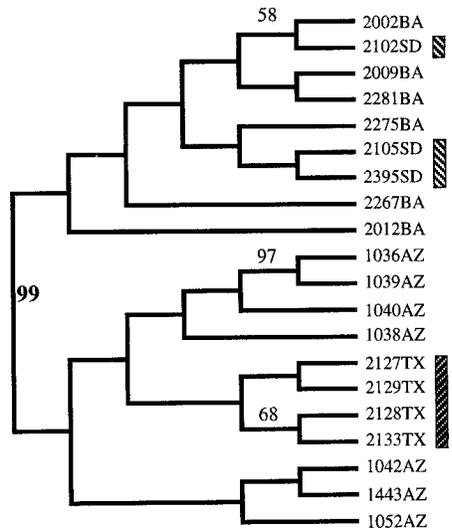


FIG. 3.—Allele distributions at seven microsatellite loci in CA (white bars;  $n = 9$ ) and non-CA (black bars;  $n = 11$ ) *Coccidioides immitis* for (A) the numbers of dinucleotide repeats found at a locus and (B) the length of the complete locus including dinucleotide repeats and flanking sequence.

A. Gene-sequence distance tree



B. Microsatellite  $D_{AS}$  tree



C. Microsatellite  $D_I$  tree

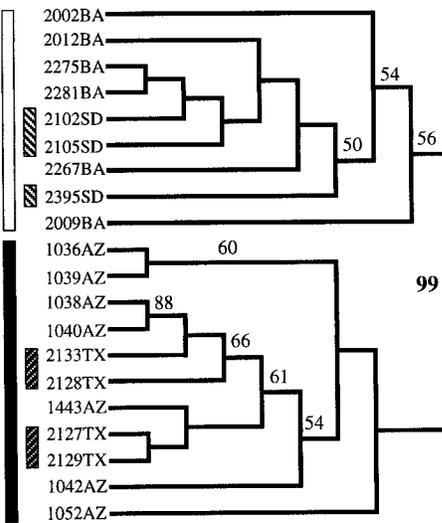


FIG. 4.—A, Neighbor-joining (NJ) tree produced by distance analysis of the combined flanking sequence and restriction fragment length polymorphism (RFLP) data sets (gene sequence data set). B, NJ tree for the microsatellite distance  $D_{AS}$ . C, NJ tree for the microsatellite distance  $D_I$ . Bootstrap values  $>50\%$  are shown as numbers above branches; numbers in bold are values for nodes separating the two species. Filled and open bars signify the *C. immitis* cryptic species CA and non-CA; hatched bars signify geographically defined populations within each species. All trees are midpoint rooted.

tween CA and non-CA occurred some  $12.8 \times 10^{-6}$  generations ago (SE =  $8 \times 10^{-6}$  generations; table 4), or 12.8 Myr if the generation time for *C. immitis* is taken as 1 year. An estimate of the time of divergence of CA and non-CA based on  $(\delta\mu)^2$ , 760,000 generations ( $\pm 280,000$ ), was an order of magnitude less than that expected and shows that at this scale,  $(\delta\mu)^2$  has not increased linearly with time. However, estimates of the time of divergence between the San Antonio (TX) and Tucson (AZ) genotypes ( $\approx 40,000$  years) were the same for both flanking-sequence and microsatellite data sets,

indicating that, at this reduced temporal scale at least,  $(\delta\mu)^2$  is approximately linear (table 4).

Discussion

We have shown that microsatellite loci may be used to resolve the population structure of *C. immitis* for both recently diverged ( $\approx 40,000$  years) and ancient ( $\approx 12$  Myr) groups by comparing microsatellite genetic distances against a multilocus organismal phylogeny inferred from flanking-sequence and RFLP loci. This ap-

**Table 4**  
**A Comparison of Lineage Divergence Times for**  
***Coccidioides immitis* Calculated from the Polymorphisms**  
**Within the Microsatellite-Flanking Sequence and  $(\delta\mu)^2$**

Pairwise Comparison	Distance Measurement	$\tau_{\text{gen}}^a$	SE
CA/non-CA	Flanking sequence	$12.8 \times 10^6$	$8.0 \times 10^6$
	$(\delta\mu)^2$	$7.6 \times 10^5$	$2.8 \times 10^5$
Non-CA/TX	Flanking sequence	$4.7 \times 10^4$	$11.6 \times 10^4$
	$(\delta\mu)^2$	$3.8 \times 10^4$	$2.3 \times 10^4$

<sup>a</sup> Coalescence time between populations in numbers of generations.

proach demonstrates the utility of microsatellites as phylogenetic, as well as population genetic, markers in this system. However, several important points have emerged from this work. (1) Genetic distances calculated from microsatellites underestimate the lengths of long branches. (2) Flanking-sequence indels add substantial phylogenetic noise to the data set. (3) Genetic distances based on the SMM perform less well than simple allele frequency data.

Empirical studies of several animal and plant taxa have shown that there are serious problems in using microsatellites to reconstruct known species relationships (Bowcock et al. 1994; Ortí, Pearse, and Avise 1997; Paetkau et al. 1997; Doyle et al. 1998). On the other hand, studies of the *Drosophila melanogaster* species complex appear to be more successful, with congruent relationships between the alcohol dehydrogenase gene and microsatellite loci being found for species separated by over 2 Myr (Harr et al. 1998). For shorter periods ( $\approx 16,000$  years), microsatellite phylogenies accurately mirror the differentiation of gray fox populations throughout the Californian Channel Islands (Goldstein et al. 1999), a result that corroborates those described here. Furthermore, the use of distances based on the SMM showed superior ability in resolving the human/chimp/gorilla clade (Goldstein et al. 1995b) compared with simple frequency-based statistics, suggesting that the development of more complex mutational models can prove useful in increasing the range of microsatellite loci for detecting species level relationships, although this was not the case with *C. immitis*.

Central to the issue of whether microsatellites make good genealogical markers is consideration of their mode and rate of evolution. Despite theory showing that distances based on the SMM will maintain linearity over millions of years (Goldstein et al. 1995a), it is apparent that for *C. immitis* these statistics are reaching a plateau relatively rapidly. Alleles at several loci (*GAC2*, *621*, *KO3*) are invariant in one or the other *C. immitis* species, indicating that constraints on free variation exist. These data from *C. immitis* are consistent with other studies showing that microsatellite distances do not increase linearly with time (Bowcock et al. 1994; Garza, Slatkin, and Freimer 1995; Lehmann, Hawley, and Collins 1996). This theory gains support from observations that few microsatellites have alleles longer than 60 repeats (Goldstein and Pollock 1997) and that there is a mutation bias to shorter alleles observed in *Drosophila*

(Schlötterer et al. 1998) and yeast (Wierdl, Dominska, and Petes 1997), as well as a loss of variability in shorter alleles (Weber 1990). As argued by Nauta and Weissing (1996), constraints in allele size, coupled with high rates of mutation, may overwhelm the effects of genetic drift, limiting the potential for genetic divergence of populations at these loci. Such effects would be expected to be especially strong in populations of microorganisms where population sizes tend to be large and generation times short. However, this does not appear to occur in *C. immitis*. Graphing the numbers of repeats for our loci demonstrated that six of the seven loci had species-specific allele distributions. Furthermore, the microsatellite distance based solely on allele frequency,  $D_{AS}$ , worked well in comparison with the organismal phylogeny. This observation shows that microsatellite allele distributions in *C. immitis* may be constrained, but not to the degree that homogenization of allele distributions occurs. As a result, species, as well as intraspecific phylogeographic groups, are readily differentiated. A single locus, *KO7*, had accumulated little genetic distance, and the use of only this locus would have led to incorrect phylogenetic conclusions. It is apparent that variation exists between loci in their patterns of divergence and diversity, and this may reflect arguments that the rate and direction of mutation vary among loci and in closely related species (Rubinsztein et al. 1995; Amos et al. 1996; but see Ellegren et al. 1997). However, taking a consensus approach corrects for the bias introduced by certain anomalous loci. The generality of this result needs to be confirmed in other species of fungi and microbes, as effective population size as well as recombination and mutation rates is expected to vary across phyla.

We show that a measure based on the proportion of alleles shared between individuals,  $D_{AS}$ , assigns *C. immitis* isolates to the correct taxonomic and phylogeographic units better than D1. This difference is due chiefly to the inability of D1 to resolve closely related lineages (San Antonio from Tucson) and is a feature associated with the high variance of this statistic (Goldstein et al. 1995a). One reason that  $D_{AS}$  behaves better than D1 is that the squaring of differences in allele size exacerbates the effect of large, multistep mutations. Alleles that are outliers in terms of size are seen at loci *ACJ* and *KO3*, and flanking-sequence variation also contributes to the creation of large alleles (locus *GAI*, CA allele 264 bp; locus *621*, CA alleles 414 and 416 bp). Mutation models have been developed to account for such multistep mutations, as well as stepwise increments in allele size (Di Rienzo et al. 1994), and may be appropriate for use here. However, in our hands a simple allele-sharing distance works well and may be sufficient if all that is needed is a method of differentiating species without specific reference to branch length. Therefore, if one's aim is to assign individuals to populations, then  $D_{AS}$  appears to be the correct measure to use due to its greater relative precision. However, if correct branching order in deeper phylogenies is desirable, then distances based on the SMM may be more appropriate, and corrections may then be used to account for constraints to free variation (Pollock et al. 1998; Zhivotovsky 1999).

Flanking-sequence indels have significant effects on the accuracy of microsatellite genealogies. Microsatellite alleles that are identical in number of repeats may have accrued length variations in flanking sequence that inflate estimates of genetic diversity (e.g., loci *GAI* and *62I*) or mask it (locus *KO7*). This is an effect that has been described in other systems (Grimaldi and Crouau-Roy 1997; Ortí, Pearse, and Avise 1997; Colson and Goldstein 1999). Sequencing a number of alleles will enable workers to choose loci that do not contain indel-prone (T)<sub>n</sub> and (G)<sub>n</sub> motifs within the flanking sequence. This approach will also enable the worker to choose loci that conform closely to the SMM and exclude loci evolving with complex mutations, enabling the use of the more sophisticated mutation models in the inference of genealogical relationships.

Our findings are important for studies in which it is not possible to assign individuals to populations before genetic analysis is performed. We are concerned about cases in which epidemics may be due to either the emergence of a novel pathogen/genotype or environmental factors magnifying a pathogen's effect. These questions are faced in recent epidemics of amphibian chytridiomycosis (Berger et al. 1998), aspergillosis of sea-fan corals (Geiser et al. 1998), and coccidioidomycosis (Fisher et al. 2000). In all cases, it is not obvious whether opportunistic pathogens are infecting environmentally stressed hosts or emerging pathogens are sweeping through unexposed populations (Morell 1999). Our results suggest that microsatellites would perform a dual purpose here, enabling cryptic population structure to be observed as well as characterizing intraspecific relationships, thus allowing the simultaneous testing of these two hypotheses.

### Acknowledgments

We thank the National Institutes of Health (NHLBI, NIAID) for financial support and two anonymous reviewers for helpful comments.

### LITERATURE CITED

- AMOS, W., S. J. SAWCER, R. W. FEAKES, and D. C. RUBINSZSTEIN. 1996. Microsatellites show mutational bias and heterozygote instability. *Nat. Genet.* **13**:390–391.
- AVISE, J. C., and R. M. BALL. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Pp. 45–67 in D. FUTUYAMA and J. ANTONOVICS, eds. *Oxford surveys in evolutionary biology*. Vol. 7. Oxford University Press, Oxford, England.
- BERGER, L., R. SPEARE, P. DASZAK, and D. GREEN. 1998. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc. Natl. Acad. Sci. USA* **95**:9031–9036.
- BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, and J. R. KIDD. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**:455–457.
- BURT, A., D. A. CARTER, G. L. KOENIG, T. J. WHITE, and J. W. TAYLOR. 1995. A safe method of extracting DNA from *Coccidioides immitis*. *Fungal Genet. Newsl.* **42**:23.
- BURT, A., B. M. DECHAIRIO, G. L. KOENIG, D. A. CARTER, T. J. WHITE, and J. W. TAYLOR. 1997. Molecular markers reveal differentiation among isolates of *Coccidioides immitis* from California, Arizona and Texas. *Mol. Ecol.* **6**:781–786.
- COLSON, I., and D. B. GOLDSTEIN. 1999. Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **152**:617–627.
- DALLAS, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome* **3**:452–456.
- DI RIENZO, 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.
- DOYLE, J. J., M. MORGANTE, S. V. TINGEY, and W. POWELL. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine* subgenus *Glycine*). *Mol. Biol. Evol.* **15**:215–218.
- ELLEGREN, H., S. MOORE, N. ROBINSON, K. BYRNE, W. WARD, and B. C. SHELDON. 1997. Microsatellite evolution: a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol. Biol. Evol.* **14**:854–860.
- FARRIS, J. S., M. KALLERSJO, A. G. KLUGE, and C. BULT. 1995. Constructing a significance test for incongruence. *Cladistics* **44**:570–572.
- FELSENSTEIN, J. 1991. PHYLIP. version 3.57c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FISHER, M. C., G. L. KOENIG, T. J. WHITE, and J. W. TAYLOR. 1999. Primers for genotyping single nucleotide polymorphisms and microsatellites in the pathogenic fungus *Coccidioides immitis*. *Mol. Ecol.* **8**:1082–1084.
- . 2000. Pathogenic clones or environmentally determined population expansion? A molecular and epidemiological analysis of an epidemic of the human pathogenic fungus *Coccidioides immitis*. *J. Clin. Microbiol.* **38**:807–813.
- 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.
- GEISER, D. M., J. W. TAYLOR, K. B. RITCHIE, and G. W. SMITH. 1998. Cause of sea fan death in the West Indies. *Nature* **394**:137–138.
- 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.
- GOLDSTEIN, D. B., G. W. ROEMER, D. A. SMITH, D. E. REICH, A. BERGMAN, and R. K. WAYNE. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**:797–801.
- GOLDSTEIN, D. B., A. RUIZ LINARES, L. L. CAVALLI-SFORZA, and M. W. FELDMAN. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**:463–471.
- . 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92**:6723–6727.
- 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. SCHLÖETTERER. 1998. A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex. *Curr. Biol.* **8**:1183–1186.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.

- HUELSENBECK, J. P., J. J. BULL, and C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. *TREE* **11**:152–158.
- KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the Hominoidea. *J. Mol. Evol.* **29**:170–179.
- KOUFOPANOU, V., A. BURT, and J. W. TAYLOR. 1997. Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **94**:5478–5482.
- . 1998. Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **95**:8414.
- 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.
- LEHMANN, T., W. A. HAWLEY, and F. H. COLLINS. 1996. An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics* **144**:1155–1163.
- LEVINSON, G., and G. A. GUTMAN. 1987a. High frequency of short frameshifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**:5322–5338.
- . 1987b. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203–224.
- LI, W., and D. GRAUR. 1991. *Fundamentals of molecular evolution*. Sinauer, Sunderland, Mass.
- MINCH, E., A. RUIZ-LINARES, D. GOLDSTEIN, M. FELDMAN, and L. L. CAVALLI-SFORZA. 1995. *Microsat* (version 1.4d): a computer program for calculating various statistics on microsatellite data. Stanford University, Stanford, Calif.
- MORELL, V. 1999. Are pathogens felling frogs? *Science* **5415**:728–731.
- NAUTA, M. J., and F. J. WEISSING. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**:1021–1032.
- OHTA, T., and M. KIMURA. 1973. The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet. Res.* **22**:201–204.
- 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.
- PAETKAU, D., L. P. WAITS, P. L. CLARKSON, L. CRAIGHEAD, and C. STROBECK. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. *Genetics* **147**:1943–1957.
- 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**:256–271.
- RUBINSZSTEIN, D. C., W. AMOS, J. LEGGO, S. GOODBURN, J. SANJEEV, L. SHI-HUA, R. L. MARGOLIS, C. A. ROSS, and M. A. FERGUSON-SMITH. 1995. Microsatellite evolution—evidence for directionality and variation in rate between species. *Nat. Genet.* **10**:337–343.
- 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.
- SHRIVER, M. D., L. JIN, E. BOERWINKLE, R. DEKA, R. E. FERRELL, and R. CHAKRABORTY. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* **12**:914–920.
- SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**:457–462.
- STEPHENS, J. C., D. A. GILBERT, N. YUHKI, and S. J. O'BRIEN. 1992. Estimation of heterozygosity for single-probe multi-locus DNA fingerprints. *Mol. Biol. Evol.* **9**:729–743.
- SWOFFORD, D. L. 1998. *PAUP\** 4.0b2a. Phylogenetic analysis using parsimony (\*and other methods). Version 4. Sinauer, Sunderland, Mass.
- TAKEZAKI, N., and M. NEI. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**:389–399.
- TAYLOR, J. W., D. M. GEISER, A. BURT, and V. KOUFOPANOU. 1999. The evolutionary biology and population genetics underlying fungal strain typing. *Clin. Microbiol. Rev.* **12**:126–146.
- TEMPLETON, A. R. 1989. The meaning of species and speciation. Pp. 3–27 in D. OTTE and J. A. ENDLER, eds. *Speciation and its consequences. A genetic perspective*. Sinauer, Sunderland, Mass.
- WEBER, J. L. 1990. Informativeness of human (dC-dT)<sub>n</sub>(dG-dA)<sub>n</sub> polymorphisms. *Genomics* **7**:524–530.
- WEBER, J. L., 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.
- ZHIVOTOVSKY, L. A. 1999. A new genetic distance with application to constrained variation at microsatellite loci. *Mol. Biol. Evol.* **16**:467–471.

ELIZABETH KELLOGG, reviewing editor

Accepted March 1, 2000