Polymorphism at the Ribosomal DNA Spacers and Its Relation to Breeding Structure of the Widespread Mushroom *Schizophyllum commune*

Timothy Y. James, Jean-Marc Moncalvo, Sean Li[1] and Rytas Vilgalys

Department of Botany, Duke University, Durham, North Carolina 27708-0338

Manuscript received May 15, 2000
Accepted for publication October 6, 2000

ABSTRACT

The common split-gilled mushroom *Schizophyllum commune* is found throughout the world on woody substrates. This study addresses the dispersal and population structure of this fungal species by studying the phylogeny and evolutionary dynamics of ribosomal DNA (rDNA) spacer regions. Extensive sampling (n = 195) of sequences of the intergenic spacer region (IGS1) revealed a large number of unique haplotypes (n = 143). The phylogeny of these IGS1 sequences revealed strong geographic patterns and supported three evolutionarily distinct lineages within the global population. The same three geographic lineages were found in phylogenetic analysis of both other rDNA spacer regions (IGS2 and ITS). However, nested clade analysis of the IGS1 phylogeny suggested the population structure of *S. commune* has undergone recent changes, such as a long distance colonization of western North America from Europe as well as a recent range expansion in the Caribbean. Among all spacer regions, variation in length and nucleotide sequence was observed between but not within the tandem rDNA repeats (arrays). This pattern is consistent with strong within-array and weak among-array homogenizing forces. We present evidence for the suppression of recombination between rDNA arrays on homologous chromosomes that may account for this pattern of concerted evolution.

UNDERSTANDING genes that are important in determining a species' distribution is a principal goal of ecological genetics. A related question asks how a gene is distributed throughout a species' range. With this latter information in hand, it becomes possible to speculate how life history traits influence the spatial distribution of a gene. We have applied this approach in studying the distribution of genes throughout the range of the cosmopolitan mushroom, *Schizophyllum commune* Fr. Species with cosmopolitan distributions are atypical, and understanding what genetic and life history traits are important in maintaining the global success of these species may suggest basic principles in understanding other species' distributions. Our primary question has been to resolve whether or not long-distance spore dispersal is important in determining the population genetics and widespread distribution of this common fungus.

S. commune is a model organism of the class Basidiomycota that includes the mushroom-forming, rust, and smut fungi. Most research on *S. commune* has focused on its complex mating system in which haploid individuals must be heteroallelic at two unlinked factors (the A and B incompatibility loci) for mating to occur. Interpopulation compatibility tests have demonstrated that *S. commune* is a single biological species (Raper *et al.* 1958). The species has been found as a wood decomposer of over 150 genera of flowering plants (Cooke 1961) and is well documented as a wound parasite of trees and as a human pathogen of minor but increasing importance (Rihs *et al.* 1996).

The global distribution of several *S. commune* genes has been previously determined. The first genes studied were the A and B incompatibility loci. Through extensive genetic crosses, Raper *et al.* (1958) determined that the numerous mating types (alleles) were randomly distributed throughout the world. The random distribution of mating types in *S. commune* was generally accepted as the result of widespread dispersal of *S. commune* spores. Although long-distance airborne dispersal of fungal spores is well documented in literature (Pady and Kapica 1955; Hirst and Hurst 1967; Ingold 1971), such dispersal may not be necessary to explain the mating type distributions of *S. commune*. This is primarily because incompatibility loci in fungi are strongly influenced by a form of negative frequency-dependent selection that selects for rare mating types and tends to maintain a large number of mating types at near-equal frequencies (May *et al.* 1999). This same selective pressure tends to equilibrate allele frequencies among populations (Wright 1939; Zambino *et al.* 1997).

Recently, results from allozyme markers (James *et al.* 1999) have challenged the notion that long-distance dispersal affects the population genetic structure of *S. commune*. These results demonstrated that most geographically disjunct populations are genetically near-

Corresponding author: Tim James, Department of Botany, Box 90338, Duke University, Durham, NC 27708-0338. E-mail: tyj2@duke.edu

[1] *Present address:* Department of Biochemistry, University of Illinois, Urbana, IL 61801.

isolated and hence experience little exchange through spores. Allelic data such as those derived from allozymes are often used to infer patterns of population structure and gene flow among samples. One drawback from such an approach is that historical processes, such as founder or vicariant events, can be difficult to disentangle from patterns resulting from a current genetic drift/gene flow equilibrium (Felsenstein 1982; Neigel 1997). As an alternative to protein electrophoretic data, DNA sequence variation has the advantage that the history of the alleles or haplotypes can be inferred through phylogenetic methods. By knowing the phylogenetic history of the alleles at a locus, there is promise of separating the past demographic events among a group of populations from current patterns of migration (Neigel 1997; Templeton 1998).

In the fungi, the ribosomal RNA coding cistron (rDNA) has been widely utilized for molecular systematic studies (*e.g.*, White *et al.* 1990; Moncalvo *et al.* 2000) as well as population genetics (*e.g.*, Ramsfield *et al.* 1996; Carbone *et al.* 1999). At the rDNA locus the amount of variation among copies within a single species is often quite dramatically lower than that between species, a pattern termed concerted evolution (Dover 1982; Arnheim 1983). In *S. commune*, the rDNA is composed of ~120 tandemly repeated units on the eighth largest chromosome (Dons and Wessels 1980; Ásgeirsdóttir *et al.* 1994). As in many other fungi, the 5S rDNA gene is also present within the repeated rDNA unit and is found between the 25S and 18S genes (Figure 1A). Two noncoding regions exist in each repeat, the internal transcribed spacer (ITS) and the intergenic spacer (IGS).

This study describes rDNA spacer variation in *S. commune* throughout the species' entire geographic range. We explore the species' population genetic structure and compare the rDNA data with the data that exist for other loci, namely the population structure of the mating incompatibility and allozyme loci. This study builds on our understanding of patterns of gene flow in *S. commune* by employing a larger and more diverse sample and by using a phylogenetic approach that has the potential of separating population structure from population history. These data also illuminate how patterns of concerted evolution progress in isolated populations that have not undergone speciation.

MATERIALS AND METHODS

Sampling strategy: A subset of 80 strains from the global *S. commune* collection used by Raper *et al.* (1958) was provided courtesy of R. C. Ullrich (University of Vermont). This sample consisted of 68 homokaryotic (genetically haploid) and 12 dikaryotic (genetically diploid) strains. Additional fruiting bodies were collected from natural substrates, primarily during the years 1995–1999, and were used to derive 98 mycelial cultures from germinated spores. Last, 16 fruiting bodies of diverse geographic origin were obtained as specimens from the herbaria at Uppsala (UPS), Sweden and Leiden, The Netherlands (Rijksherbarium). Details on the origin of strains and fruiting body collections are available from the authors.

DNA techniques: Tissue for DNA extraction was obtained from *S. commune* strains by scraping mycelia from week-old cultures grown on either potato dextrose agar (Difco, Detroit) or a medium consisting of malt extract (1.5%), yeast extract (0.3%), glucose (0.5%), and agar (1.5%). Mycelia were dehydrated in a Speed-vac concentrator (Savant Instruments, Farmingdale, NY) and ground to a fine powder using a dissecting needle and microcentrifuge pestle. Total genomic DNA was isolated from the ground mycelia following Zolan and Pukkila (1986). Herbarium specimens were extracted in a similar manner using ~0.1 g of fruiting body tissue.

The structure of the rDNA locus in *S. commune* and the location and orientation of primers for PCR are shown graphically in Figure 1A, with unpublished primer sequences provided in the legend. The primers for the ITS region (ITS1 and ITS4) are reported in White *et al.* (1990). For the IGS1 locus, the primers LR20R and 5SRNA were used for PCR amplification, and internal primers LR12R and 5SRNASC were employed for automated sequencing. Primers annealing to the 5S and 18S rDNA genes (5SRNAR and INVSR1R) were used to amplify the IGS2 region, and these, as well as internal primers (IGS2R1-4), were used to obtain nearly complete IGS2 sequences (~2.5 kbp).

Amplification of the ITS and IGS1 loci was performed using reaction conditions and thermocycling parameters described in Vilgalys and Hester (1990). For the IGS2 locus, amplification cycles were modified such that annealing was at 57° and extension for 75 sec at 72°. Amplicons were purified with ULTRAfree-MC centrifugal columns (Millipore Corp., Beford, MA) and sequenced using dye terminator chemistries on ABI 373 or 377 DNA sequencers (Perkin Elmer, Norwalk, CT). Restriction analysis of amplicons was conducted using the enzyme *Hin*fI in a total reaction volume of 15 μl following the manufacturer's protocol (Promega, Madison, WI).

To separate different IGS1 haplotypes within heterogeneous PCR reactions, amplicons from both dikaryons and fruiting bodies were cloned into the PCR2.1 vector (Invitrogen, San Diego). A single recombinant bacterial colony was chosen for each sample and placed directly into a 25-μl volume PCR reaction lacking any template. The primers LR20R and 5SRNA were used to reamplify and sequence these clones from PCR2.1.

Sequences for IGS1 are deposited in GenBank under accession nos. AF249392–AF249586. Accession numbers for ITS are AF249364–AF249391, and numbers for IGS2 are AF249355–AF249363.

Phylogenetic analyses: Sequences were aligned by eye using the program GeneDoc (Nicholas and Nicholas 1997) and imported into the phylogenetic software package PAUP* (Swofford 1998). For the IGS1 locus, the 195 sequences were used to generate an alignment of 671 bp. An alignment of 626 bp was produced for 27 sequences from the ITS region, and the IGS2 data set employed nine samples aligned into a 2704 bp matrix. At IGS1, sequence alignment was mostly unambiguous; however, several regions were difficult to align and were excluded for the purpose of phylogenetic analysis, leaving 506 aligned characters. Many gaps existed in the alignment of IGS1 and were considered to be an additional character state for phylogenetic analyses. When gaps were greater than a single nucleotide long, these gaps were recoded such that each apparently unique insertion/deletion event was considered a separate character state. For the ITS, gaps in the alignment were also considered a fifth character state, whereas gapped regions were excluded from the analysis of IGS2 due to the uncertainty in the alignment of the repetitive elements.

Phylogenies were inferred using the maximum parsimony criterion. For IGS1 and ITS regions heuristic searches were performed to find the most parsimonious phylogenies, employing random sequence addition to find initial trees and using the tree-bisection and reconnection (TBR) algorithm for branch swapping (Swofford 1993). For the IGS1 data, 100 heuristic searches were performed, saving at most 100 most parsimonious trees per search, and for the ITS region 1000 searches were performed, saving all most parsimonious trees. Exhaustive searching of all possible topologies was used to find most parsimonious trees for IGS2. Bootstrap confidence measures for phylogenetic nodes were generated using 100 replicate resamplings, each subject to a heuristic search with TBR branch swapping. For ITS and IGS2 data sets, all most parsimonious trees were saved for each bootstrap replicate. For the IGS1 data set, a maximum of 50 trees were saved for each replicate. Homoplasy (convergent or parallel evolution) was measured on the molecular phylogenies through the consistency index (CI) statistic (Farris 1989). This statistic provides a ratio of the minimum number of character changes along a phylogeny given the data to the actual number of changes along the phylogeny as reconstructed by the parsimony criterion.

Measures of population structure and nucleotide diversity: Nucleotide diversities, or the average number of differences per site between two homologous sequences (π) were calculated using the program DNASP (Rozas and Rozas 1997) following equation 10.5 of Nei (1988). Diversities were calculated both for the global population as a whole as well as separately for the three major clades (see RESULTS). Coding and gapped regions were excluded from the diversity computations.

Wright's F_{ST} was used to describe what proportion of the total genetic variance at a locus is due to differences among populations (Wright 1951). Sequences were grouped by sample origin into eight populations: Africa, Asia, Australasia, the Caribbean, Central America, Europe, North America, and South America. These populations are nearly the same as those in previous studies (Raper *et al.* 1958; James *et al.* 1999). F_{ST} can range from 1.0, in which all of the variation is among populations, to 0.0, in which all populations appear homogeneous. We calculated F_{ST} via an AMOVA analysis using the software package ARLEQUIN (Excoffier *et al.* 1992; Schneider *et al.* 1997).

Interpreting analyses of molecular variation (*e.g.*, F_{ST}) as a reflection of gene flow relies on the assumption that associations of haplotypes with geography result from a gene flow/genetic drift equilibrium among populations. Alternatively, nested clade analyses can be used to discriminate between phylogeographic patterns resulting from historical events (*e.g.*, past fragmentation, range expansion, colonization) and those due to recurrent gene flow (Templeton *et al.* 1995; Templeton 1998). These analyses utilize a phylogeny and the geographical distribution of haplotypes or clades of haplotypes. Nested clade analysis (NCA) was performed using the software GEODIS v. 2.0 (Posada *et al.* 2000) on subsets of the IGS1 network (phylogeny) for which there was a low probability ($P < 0.05$) of containing any convergent mutations (described in Templeton *et al.* 1992). Nested clade analysis divides a phylogeny into hierarchical subsets of *n* step clades, where *n* is an integer equal to the number of observed mutations connecting the haplotypes of a clade. For each clade there is a geographic focal point about which all individuals are centered. Two quantitative measures of how haplotypes within a clade are geographically dispersed were suggested by Templeton *et al.* (1995). D_C (X) describes the average distance of each haplotype within clade X from the geographic center of clade X. D_n (X) similarly describes the average distance of each haplotype within clade X from the geographical center of the next more inclusive clade within which clade X is nested. Values significantly larger or smaller than that expected for no geographic association of haplotypes can be tested via permutation of population assignments to haplotypes (Templeton *et al.* 1995). Haplotypes were assigned to the same eight populations defined for F_{ST} estimation. Clade distances were calculated using a distance matrix among populations. Contiguous locations (*e.g.*, North and Central America) were assigned a distance value of one. Locations separated by Atlantic, Pacific, or Indian Oceans were assigned an arbitrary distance value of three. Although not ideal, the use of approximate values in the distance matrix instead of geographic coordinates was necessitated because population samples were collected from diverse localities within geographically broad populations.

RESULTS

Molecular variation within the spacer regions: Preliminary sampling showed the ITS region to possess considerably less sequence variation than the IGS1 region. Therefore IGS1 was selected for more intensive sampling. Of 195 strains sampled for IGS1, 143 unique haplotypes were identified. The average number of substitutions per site (π) among any two random samples from the global population was 0.044 ± 0.002. Of the 506 aligned positions at IGS1, 172 (or 33.9%) displayed nucleotide polymorphisms or polymorphic insertion/deletion events (indels). Of the variable positions, 80 (or 46.5%) of these were unique to a single sample, (*i.e.*, singletons). In contrast, of the 27 sequences obtained for ITS, only 31 positions (or 5.0%) were variable, and the corresponding value of π was 0.007 ± 0.001. For the nine IGS2 sequence samples, 252 of 2176 sites (or 11.6%) were variable, with $\pi = 0.049 \pm 0.004$. Sequence diversity was higher at the 5′ half of IGS2 ($\pi = 0.065 \pm 0.007$) relative to that of the 3′ half ($\pi = 0.040 \pm 0.004$). Moreover, only two sites were variable in the 200-bp region directly upstream of the 18S rDNA. The spectrum of polymorphism along each of the three spacer regions (measured as π along a sliding window of 25 positions; Figure 1B) contrasts the high diversity of intergenic spacers with both coding regions and the ITS.

Regions of invariant sequence within the IGS1 spacer were also uncovered. For instance, all sequences possessed a "TATA box" at nucleotides −29 to −25 relative to the 5S rRNA gene. The "TATA box" of *S. commune* is in the same position as the consensus sequence of *Laccaria bicolor* (Martin *et al.* 1999) and *Neurospora crassa* (Selker *et al.* 1986), and is found as an upstream element of the 5S rDNA in all organisms studied to date (Venkateswarlu *et al.* 1991).

In the IGS2 spacer, beginning ~40 bp downstream from the 5S rRNA gene, a large region (~400 bp in length) of highly repetitive DNA was encountered. This consisted of numerous 7-bp repeats, with the repeat TCAGTA(G/A) encountered most frequently. Between

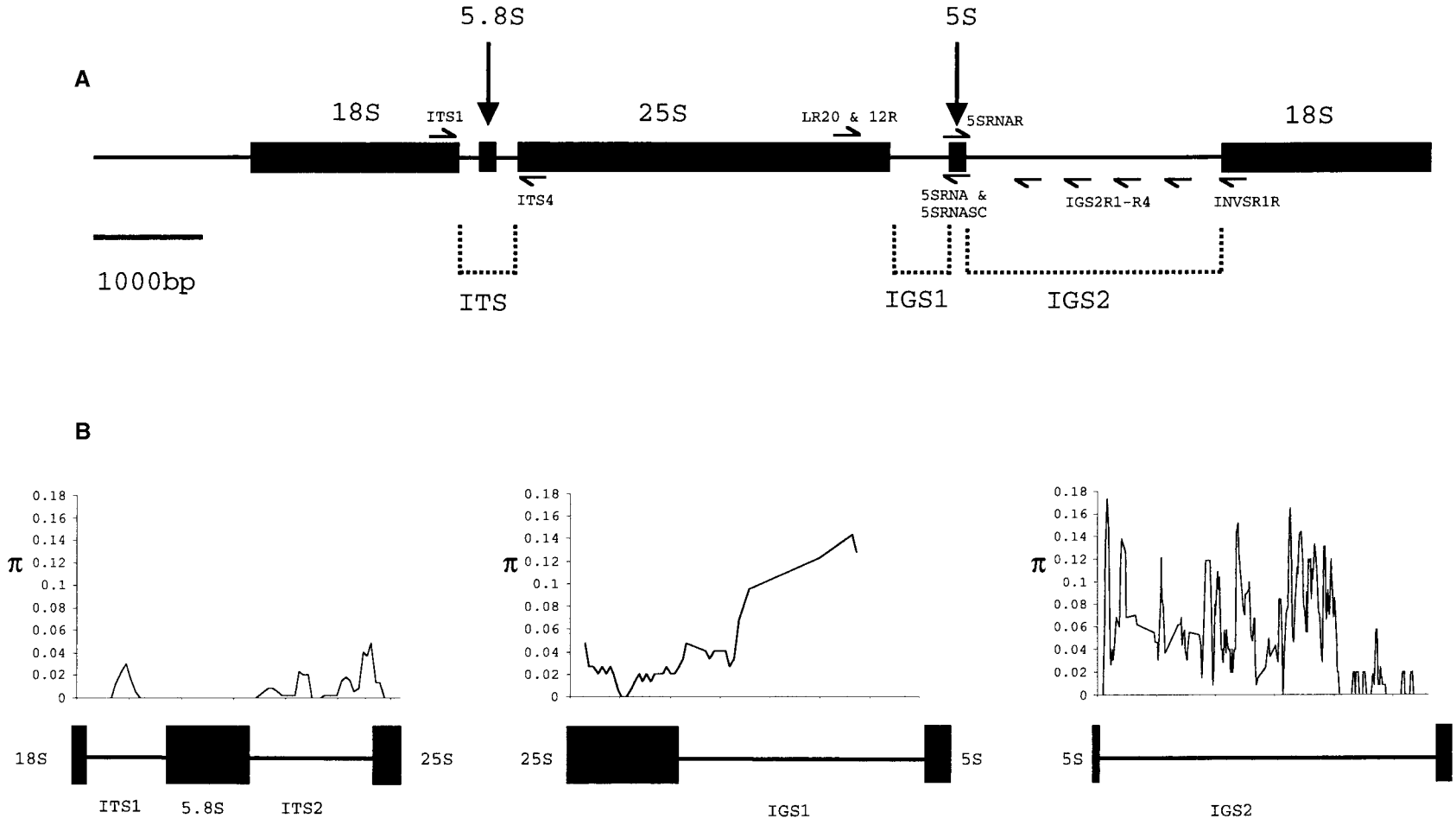


FIGURE 1.—(A) Schematic of rDNA locus in *Schizophyllum commune*. Boxes indicate coding regions. Half arrows indicate primers used for PCR and sequencing. Unpublished primer sequences were as follows: LR20R, 5′-GTGAGACAGGTTAGTTTTACCCT; LR12R, 5′-GAACGCCTCTAAGTCAGAATCC; 5SRNA, 5′-ATCAGACGGGATGCGGT; 5SRNAR, 5′-ACQGCATCCCGTCTGAT; 5SRNASC, 5′-GGGATGCGGTGCTTTC; INVSR1R, 5′-ACTGGCAGAATCAACCAGGTA; IGS2R1, 5′-AACATTGCAAGCGACCGGCAGTT; IGS2R2, 5′-GGTTTTGGCTGACTTTGATTTTG; IGS2R3, 5′-CACGAGTTTTATGGCCACTG; IGS2R4, 5′-CAGTACTAACAGTCCTTGGTA. (B) Spectrum of polymorphism among the three rDNA spacer regions depicted as a sliding window of size 25 bp and step size of 8 bp. π describes the average number of pairwise substitutions at a particular site among the global population. The abscissa represents the position along the spacer region about which the sliding window is centered.

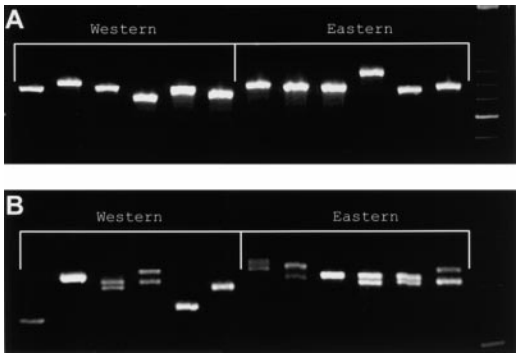


FIGURE 2.—Extreme spacer length variation near the 5′ end of IGS2 detected using PCR primers 5SRNAR and IGS2R4. Strains are grouped into geographic origin (eastern and western hemispheres). A shows monokaryotic strains that appear to have only a single length variant, whereas in B dikaryotic strains frequently possess two variants.

15 and 40 of these repeats were found in each strain interspersed among nonrepetitive DNA. Larger repetitive elements (19 bp) primarily in tandem arrangement were found near the middle of the IGS2 spacer, and 4–8 imperfect copies of this element were observed among strains. No large (*i.e.*, >50 bp) repeats at any spacer region were found by dot plot analysis.

Sequence data revealed the ITS1 region to be ~150 bp and the ITS2 region to be ~240 bp. Both ITS spacers displayed minimal among-strain length heterogeneity (<4 bp per spacer). In contrast, the IGS1 spacers demonstrated much greater length heterogeneity between strains, ranging in size from 280 to 340 bp. The size of the IGS2 region also varied between strains and was determined to be 2400–2500 bp from DNA sequencing.

Concerted evolution of rDNA spacers: Heterogeneity was observed in sequence chromatograms generated by direct sequencing of IGS1 amplicons from many dikaryotic individuals. This indicated that these genetically diploid samples possessed more than one IGS1 haplotype of differing length, causing a reading shift in the sequence chromatograms. In contrast, sequence chromatograms for monokaryotic or haploid individuals always produced distinct, unambiguous data, suggesting that all or most of the copies within these strains were of equal length. Length variation at IGS2 also appeared confined to dikaryotic individuals and absent within monokaryons. Amplification products of the IGS2 region using the primers 5SRNAR and IGS2R4 often produced double-banded products for dikaryotic individuals (Figure 2). At the ITS region, no length or sequence variation was ever observed among the amplification products of dikaryons.

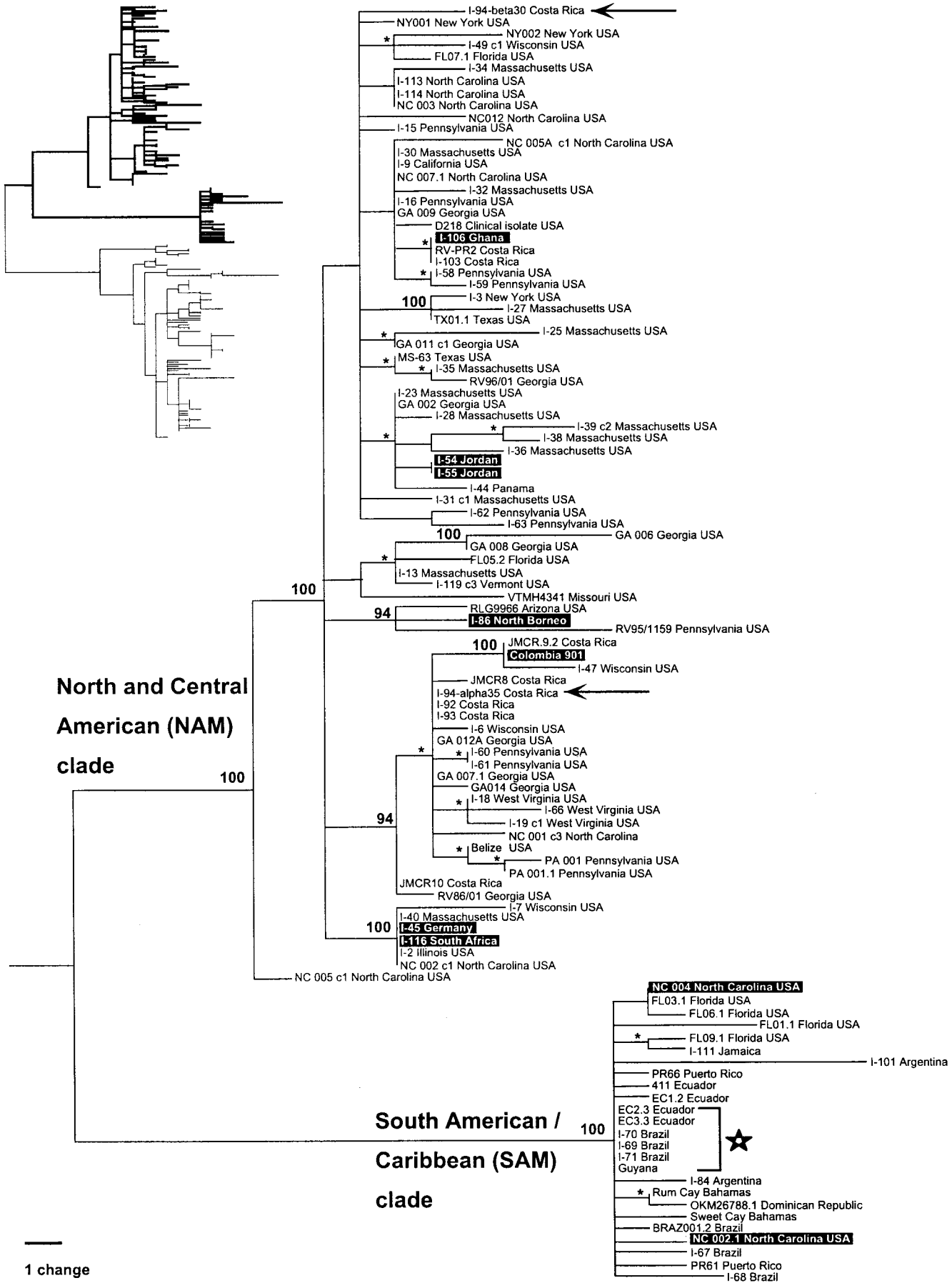
After cloning IGS1 amplicons from dikaryotic samples into PCR2.1, typically a single cloned haplotype was sequenced. For one sample in which two clones were sequenced, two different haplotypes were found (1-94-alpha and 1-94-beta Costa Rica; arrows in Figure 3).

Since sequence heterogeneity appeared largely restricted to between-array rather than within-array variation, *i.e.*, confined to different homologous chromosomes, we investigated whether recombination was suppressed within this large gene region. The cross between strains Belize#1 and Ecuador1.2 involved two monokaryons, each of whose chromosomal arrays appear homogenous at IGS1. However, Belize#1 and Ecuador1.2 arrays differ in the size of the IGS1 regions, are in two different clades in the IGS1 phylogeny, and differ in the presence of a *Hin*fI restriction site within the spacer. Progeny ($n = 74$) were isolated from this cross and were genotyped at the *Hin*fI restriction site. All progeny genotypes were uniallelic, with 41 progeny identical to the Belize#1 parental type and 33 identical to the Ecuador1.2 type. In summary, there was no evidence from this cross for recombination between ribosomal arrays on different chromosomes.

Phylogeographic patterns in spacer evolution: Phylogenetic reconstruction of the IGS1 data revealed three primary groupings, each supported with 100% bootstrap confidence. One of the groupings corresponded to samples from North America (NAM) and Central America (NAM clade; Figure 3). However, only one out of seven samples from western North America were found in this group. A second group consisted of samples from South America (SAM) and the Caribbean, and included some isolates from Florida and North Carolina (SAM clade; Figure 3). The final group contained the large majority of the samples from the eastern hemisphere (EAS clade; Figure 4). This last group also included most samples from western North America and Wisconsin. Besides the placement of the western North American samples, a lack of geographic concordance with phylogeny occurred in other instances. For example, some isolates from Germany, South Africa, and Jordan grouped within the North and Central American clade (Figure 3). These geographic discrepancies could be most easily explained by recent migration events.

Although the samples generally grouped by geographic origin, there is little evidence of geographic substructure within each of the three major groups. There were, however, two subgroups in the EAS clade that clustered by geography (Figure 4). One subgroup consisted of many samples from Australia and Papua New Guinea, and was defined by three unique changes. The other subgroup consisted of western North American strains intermixed with European samples. None of the samples from western North America were identical with those from Europe; however, some of these samples differed by only a single base change.

Phylogenetic analysis of the ITS alignment recovered the same three major geographic groupings as seen at IGS1 (Figure 5A). Due to the low level of variation among sequences, bootstrap values for these groupings were very low. Despite this, the variable sites were very


I-94-beta30 Costa Rica
NY001 New York USA
NY002 New York USA
I-49 c1 Wisconsin USA
FL07.1 Florida USA
I-34 Massachusetts USA
I-113 North Carolina USA
I-114 North Carolina USA
NC 003 North Carolina USA
NC012 North Carolina USA
I-15 Pennsylvania USA
NC 005A c1 North Carolina USA
I-30 Massachusetts USA
I-9 California USA
NC 007.1 North Carolina USA
I-32 Massachusetts USA
I-16 Pennsylvania USA
GA 009 Georgia USA
D218 Clinical isolate USA
I-106 Ghana
RV-PR2 Costa Rica
I-103 Costa Rica
I-58 Pennsylvania USA
I-59 Pennsylvania USA
100
I-3 New York USA
I-27 Massachusetts USA
TX01.1 Texas USA
I-25 Massachusetts USA
GA 011 c1 Georgia USA
MS-63 Texas USA
I-35 Massachusetts USA
RV96/01 Georgia USA
I-23 Massachusetts USA
GA 002 Georgia USA
I-28 Massachusetts USA
I-39 c2 Massachusetts USA
I-38 Massachusetts USA
I-36 Massachusetts USA
I-54 Jordan
I-55 Jordan
I-44 Panama
I-31 c1 Massachusetts USA
I-62 Pennsylvania USA
I-63 Pennsylvania USA
100
GA 006 Georgia USA
GA 008 Georgia USA
FL05.2 Florida USA
I-13 Massachusetts USA
I-119 c3 Vermont USA
100
VTMH4341 Missouri USA
94
RLG9966 Arizona USA
I-86 North Borneo
RV95/1159 Pennsylvania USA
100
JMCR.9.2 Costa Rica
Colombia 901
I-47 Wisconsin USA
JMCR8 Costa Rica
I-94-alpha35 Costa Rica
I-92 Costa Rica
I-93 Costa Rica
I-6 Wisconsin USA
GA 012A Georgia USA
I-60 Pennsylvania USA
I-61 Pennsylvania USA
GA 007.1 Georgia USA
GA014 Georgia USA
I-18 West Virginia USA
I-66 West Virginia USA
94
I-19 c1 West Virginia USA
NC 001 c3 North Carolina
Belize USA
PA 001 Pennsylvania USA
PA 001.1 Pennsylvania USA
JMCR10 Costa Rica
RV86/01 Georgia USA
I-7 Wisconsin USA
I-40 Massachusetts USA
100
I-45 Germany
I-116 South Africa
I-2 Illinois USA
NC 002 c1 North Carolina USA
NC 005 c1 North Carolina USA
North and Central American (NAM) clade
100
NC 004 North Carolina USA
FL03.1 Florida USA
FL06.1 Florida USA
FL01.1 Florida USA
FL09.1 Florida USA
I-111 Jamaica
I-101 Argentina
PR66 Puerto Rico
411 Ecuador
EC1.2 Ecuador
EC2.3 Ecuador
EC3.3 Ecuador
I-70 Brazil
I-69 Brazil
I-71 Brazil
Guyana
South American / Caribbean (SAM) clade
100
I-84 Argentina
Rum Cay Bahamas
OKM26788.1 Dominican Republic
Sweet Cay Bahamas
BRAZ001.2 Brazil
NC 002.1 North Carolina USA
I-67 Brazil
PR61 Puerto Rico
I-68 Brazil
1 change

consistent with the phylogeny, as indicated by a high CI of 0.969. The western North American sample grouped with the eastern hemisphere samples, as observed in the IGS1 tree.

Phylogeographic patterns of sequence evolution at IGS2 resembled those seen with IGS1 and ITS regions (Figure 5B). Despite the limited sampling, the same three geographic clades were well supported. The large amount of IGS2 variation also resolved a very well supported and consistent (CI = 0.964) phylogeny.

Although the three spacer regions showed topological congruence, they differed completely in the rank order of sequence diversity for each of the three major clades (also apparent in relative phylogenetic branch lengths). For each spacer phylogeny, a different clade displayed the highest levels of nucleotide polymorphism as measured by substitutions per site (π). At IGS1, the rank-order of sequence variation followed EAS ($\pi = 0.018 \pm 0.001$) > NAM ($\pi = 0.015 \pm 0.001$) > SAM ($\pi = 0.013 \pm 0.002$). The SAM clade also displayed the lowest IGS2 polymorphism, with rankings NAM ($\pi = 0.027 \pm 0.010$) > EAS ($\pi = 0.024 \pm 0.007$) > SAM ($\pi = 0.011 \pm 0.003$). Finally, a reversal of IGS1 polymorphism was observed at ITS, where SAM ($\pi = 0.008 \pm 0.010$) > NAM ($\pi = 0.006 \pm 0.002$) > EAS ($\pi = 0.003 \pm 0.001$).

Population structure and dispersal: The extensive sampling of IGS1 sequences allowed estimation of F_{ST}, or the proportion of genetic variation due to differences among populations. The estimate of global F_{ST} derived using AMOVA was 0.487, implicating strong geographic population subdivision within *S. commune*. The IGS1 F_{ST} value is higher than that estimated from allozyme data (0.214; James *et al.* 1999; Table 1) and contradicts the low level of among-population differentiation inferred from the distribution of mating types ($F_{ST} = 0.008$ and 0.010 for the A and B mating loci, respectively; Raper *et al.* 1958; Table 1).

While estimates of F_{ST} rely on equilibrium population dynamics, NCA do not and are capable of detecting nonequilibrium historical associations between populations. Two of the three major geographic clades (SAM and EAS) were subject to NCA. Within the SAM clade (see Figure 3), there is little phylogenetic structure. A parsimonious network of these sequences is unambiguous and results in many tip clades emanating from a single internal haplotype (Figure 6A). NCA shows that the internal haplotype, as well as the one-step clade it is contained within (1-12), shows significantly ($P < 0.05$) small geographic dispersion (or clade distance, D_c) relative to the distribution of D_c values generated by permuting population assignments of haplotypes (Figure 6B). In addition, since the internal clade 1-12 shows lower clade distance than the average of the clade distances for the tip clades that have presumably descended from it [*i.e.*, D_c-D_cT significantly low ($P < 0.05$)], the inference key presented in Templeton (1998) suggests a contiguous range expansion within the SAM clade. Since the internal clade was largely restricted to northern South America, it is inferred that a range expansion into the Caribbean as well as North America and Argentina has occurred from this ancestral haplotype.

For the EAS clade (Figure 4), much greater sequence variation exists and such variation causes ambiguity in network estimation. Regardless, significant historical patterns were found in this group. Clearly, because western North American haplotypes are nested within the European clade (Figure 4), a long distance colonization of America from Europe is suggested. Such a colonization is supported by NCA because the clade distances (D_c) of the clades containing western North American haplotypes are significantly low ($P < 0.05$) but their nested clade distances (D_n) are significantly high (results not shown). Such reversals between D_c and D_n for a clade generally indicate dispersal events (Templeton 1998).

DISCUSSION

The data presented here further explore the geographic scale and composition of breeding populations in *S. commune*. The sequence data show three genetically discrete populations that appear to be geographically broad. More precisely, most of the eastern hemisphere forms one phylogenetic group; most of North America, including Central America, comprises the second group; and the third group contains the majority of the South American and Caribbean samples. Although our extensive haplotype sampling has been able to reveal instances of rare migration (Figures 3 and 4), these migrations may have occurred only in the recent history of the species.

We also investigated the evolution of all rDNA spacer regions to understand how population subdivision affects the amount and pattern of sequence variation at loci evolving under different rates of nucleotide substitution. The IGS displayed high levels of nucleotide diversity in contrast to the minimal variation observed at

Figure 3.—Phylogeny based on IGS1 sequence data. Heuristic searches found 9700 equally parsimonious trees. One of the trees is shown, using parsimony to describe branch lengths (total length of the tree = 285 steps; CI = 0.795). This portion of the tree covers the western hemisphere group. Sequences are labeled with the strain or collection name, followed by geographical origin. Bootstrap values are shown above nodes; *, short internal node with >90% bootstrap support. Sequences that are placed contradictory to their geographic origin are shown as white-on-black text. The group of identical sequences indicated by a star is the set of haplotypes shown by nested clade analysis to be an internal clade with significantly small geographic dispersion (see Figure 6A).

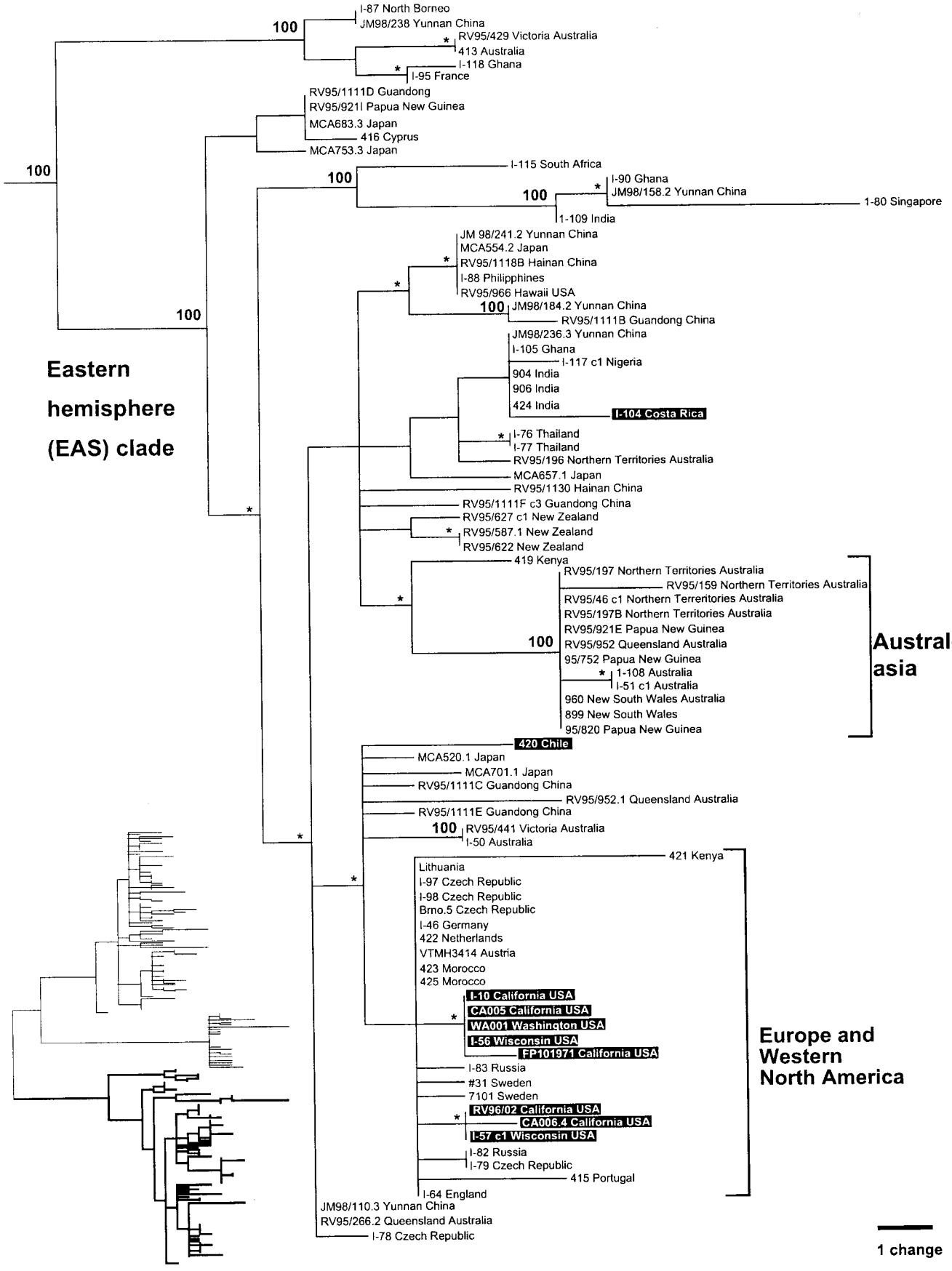


FIGURE 4.—Second portion of the IGS1 phylogeny showing the eastern hemisphere (EAS) group.

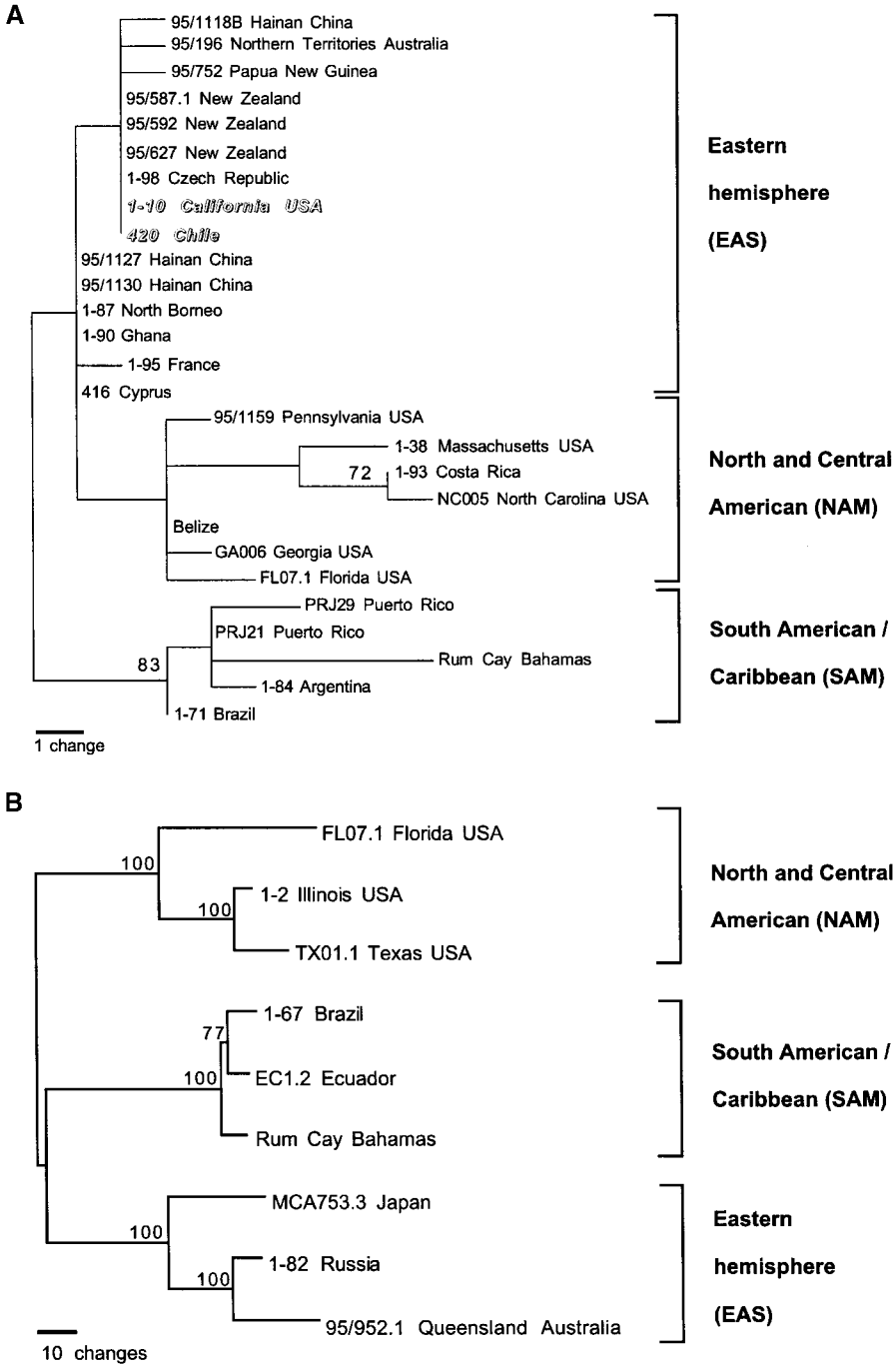


FIGURE 5.—(A) Phylogeny based on ITS sequence data. Presented is one of three most parsimonious trees, 32 steps long (CI = 0.969). (B) Phylogeny based on IGS2 sequence data. Exhaustive searching recovered one most parsimonious tree, 305 steps long (CI = 0.964).

the ITS. However, all rDNA spacer regions displayed the same three geographic groupings using phylogenetic methods.

Global distribution of genetic elements—selection *vs.* drift: *S. commune* is a common, cosmopolitan species for which neither morphology, mating compatibility tests, nor distributions of mating types have differentiated global populations. In contrast, neutral marker loci have shown that very strong genetic differentiation exists among geographic populations (Table 1). The random distribution of mating types among otherwise genetically divergent *S. commune* populations can most easily be explained as a demonstration of strong balancing selection on mating loci. This selection prevents differentiation and genetic drift of mating type frequencies despite the lack of significant interpopulation gene flow (JAMES *et al.* 1999).

Both the lack of mating type differentiation among continents and the lack of mating incompatibility among intercontinental populations in *S. commune* may have a general genetic basis. Many biological species of mushroom fungi are broadly distributed over multiple continents (HORAK 1983; VILGALYS and SUN 1994; PETERSON and MCCLENEGHAN 1997 and references

TABLE 1

F_{ST} estimates among global populations of *Schizophyllum commune* at selected loci

Data	F_{ST}
A mating-type locus	0.008
B mating-type locus	0.010
Allozyme (multilocus estimate)	0.214
Allozyme (aspartate aminotransferase)	0.476
Allozyme (aconitase)	0.442
Allozyme (sorbitol dehydrogenase)	−0.023
Allozyme (phosphoglucoisomerase)	0.001
IGS1 (DNA sequences)	0.487

therein) and display high levels of mating type diversity with little or no geographical structuring (Whitehouse 1949; Ullrich 1977). A model of mating compatibility evolution among allopatric populations of basidiomycete fungi should consider that two populations that become geographically isolated may not differentiate at mating loci due to balancing selection that slows the loss of mating types, allowing equivalent sets of mating types to be maintained between the two populations (Zambino *et al.* 1997; May *et al.* 1999). Since the mating type loci themselves additionally function in sexual development (Hiscock and Kües 1999), it is likely that populations that do not possess the same sets of mating types would show a high degree of incompatibility due to nonfunctional interactions between mating types that have evolved independently.

The patterns of phylogeography at *S. commune* IGS1 were similar to those derived from allozyme markers (James *et al.* 1999), *i.e.*, geographically broad but genetically divergent populations. In this study, the IGS1 data provide much more detail in defining populations than allozyme loci due to the ability to produce a robust phylogeny. Additionally, phylogenetic methods were capable of detecting in *S. commune* the signatures of colonization and range expansion that could not be addressed using allele frequency-based estimates of population structure (see results).

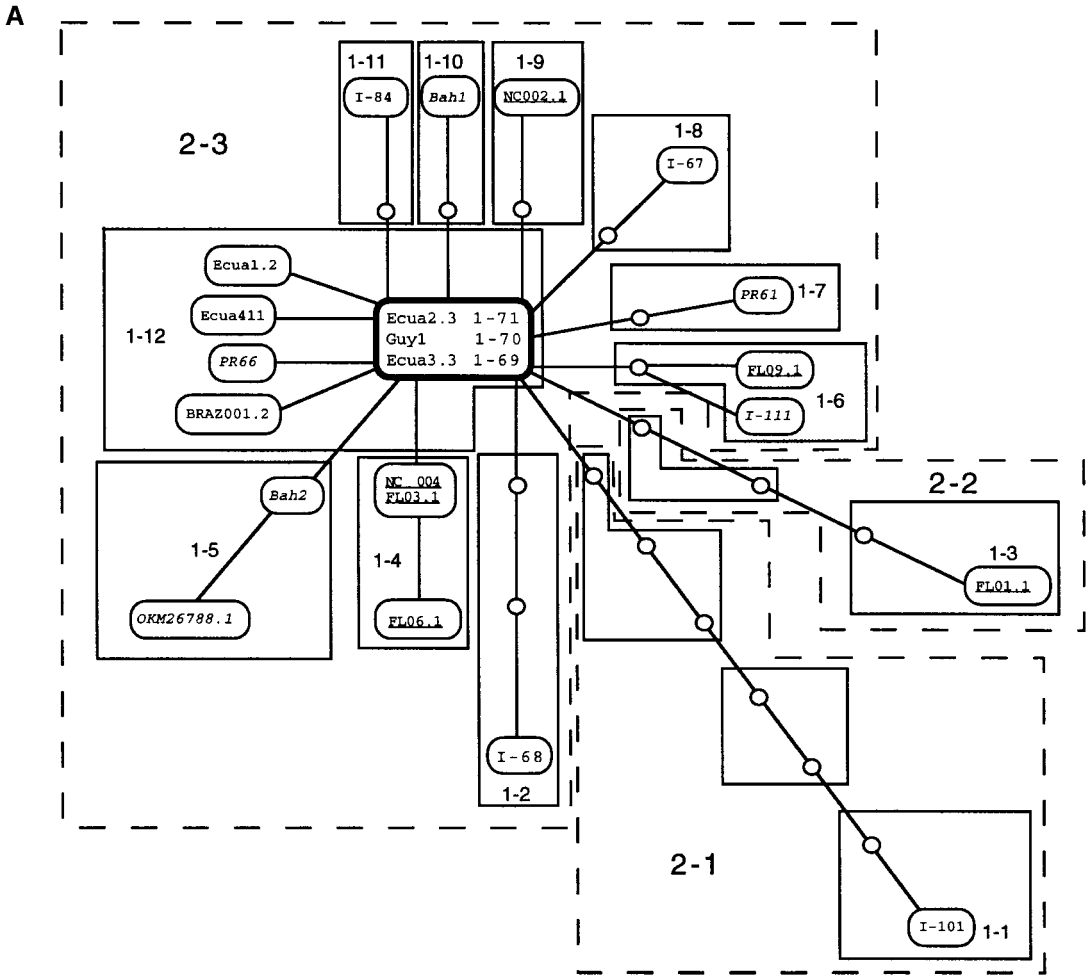
Among loci of *S. commune* that are believed to be evolving in a neutral or near-neutral fashion, not all show population differentiation, *e.g.*, sorbitol dehydrogenase (Table 1). Menadione reductase and isocitrate dehydrogenase also show this pattern of allozyme variation in *S. commune* (James *et al.* 1999), and all three loci additionally share reduced allelic diversity. Similar patterns have been observed in *Drosophila melanogaster* in which several loci possess a single common allele present in all populations and only a few additional rare alleles at low frequency (Singh and Rhomberg 1987). Strong purifying selection at a locus could be responsible for this lack of observable genetic differentiation among populations, by the elimination of new variants of suboptimal fitness.

While the patterns of phylogeography suggest a large geographical component to population subdivision, a climatological component may also be involved. For example, the fact that four of six samples from southern Florida, a subtropical climate, group with samples from the Caribbean Sea and South America in the IGS1 phylogeny (Figure 3), suggests that the similar climate and habitat may be more important in determining population structure than geographical distance. A similar example in the Mediterranean can be found in the grouping of Moroccan samples with those of European origin (Figure 4). These results suggest that *S. commune* may disperse frequently over marine areas devoid of habitat. If such long distance dispersal is possible, why then are global populations differentiated? One hypothesis appeals to a historical process in which the former distribution of *S. commune* was much narrower and disjunct, allowing the differentiation of the three major lineages. Subsequent range expansions in combination with long-distance dispersal would then create the phylogeographic patterns we have observed, namely three genetically divergent lineages whose distributions generally, but not entirely, correspond with their geographic origin. This hypothesis is supported from the results of the nested clade analysis that suggest continuous range expansion to have occurred in at least one of the three major lineages (SAM clade; Figure 6). In addition, because the SAM clade contains the least amount of sequence polymorphism at two of the three spacers, the

Figure 6.—(A) Parsimony network of IGS1 haplotypes within the South American (SAM) clade. Strain designations enclosed by rounded boxes possess the same haplotype. Solid-line boxes surround one-step clades, and dashed-line boxes show two-step clades. Numbering indicates the clade designations utilized by nested clade analysis (*e.g.*, 1-5 is the fifth one-step clade). The large internal haplotype (star in Figure 3) represented by six strains is surrounded by a boldface, rounded box. Strains from the Caribbean are shown in italic; strains from North America are underlined; the remaining strains are from South America. Small circles indicate inferred but not sampled haplotypes. (B) Clade distances (D_c) and nested clade distances (D_n) generated by GeoDis for the one-step clades nested within clade 2-3. User-defined geographic distances among populations were as follows: North America and the Caribbean were considered one distance unit apart, South America and the Caribbean were one distance unit, and North and South America were two distance units. Also shown are the differences between average clade distances for interior *vs.* tip clades ($D_cI - D_cT$) and the corresponding value for nested clade distances ($D_nI - D_nT$). Superscripts indicate significantly small (S) or large (L) distances. The chain of inference follows the key provided in the Appendix of Templeton (1998). In words, contiguous range expansion within Clade 2-3 is suggested because clade 1-12 shows significantly small geographic dispersion and it is the only interior clade (presumably ancestral) in clade 2-3.

SAM range expansion may be of a relatively recent origin. Further support for the nonequilibrium status of *S. commune* populations exists in the high number of long-distance dispersal events recorded in the IGS1 phylogeny (Figures 3 and 4). Such migrations appear to be more frequent, or more recent, than the restricted dispersal needed to create the deep phylogenetic divergences between the three major geographic clades. If the range expansion to cosmopolitan has been rather recent, then the process of population homogenization

A



B

Clade 2-3

1-Step clades	1-2	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12
D_C	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.33^S
D_n	0.73	1.43^L	0.79	1.07	0.77	0.72	1.36	0.77	0.73	0.88

D_CI-D_CT	0.180
D_nI-D_nT	-0.132^S

Chain of inference-
1-2-11-12 NO: contiguous range expansion

through gene flow may yet be incomplete. Our observations that *S. commune* thrives in secondary forest growth and human-disturbed sites suggest that hypotheses considering range expansion, and possibly intercontinental human-mediated dispersal, merit further attention.

Concerted evolution of *S. commune* rDNA spacers: The large level of haplotype diversity at the ribosomal spacers appears to be entirely due to differences between individuals rather than differences among the tandem repeats within an array. Because haploid strains appear fixed for a single haplotype, yet many haplotypes exist within populations, a strong argument is made for the occurrence of intrachromosomal homogenization of elements in a single array and for a lack of recombination between the rDNA arrays of homologous chromosomes at meiosis. Our failure to detect recombinants between rDNA arrays in one test cross provides additional support for recombination suppression. In addition, the high consistency of the IGS1 phylogeny (CI = 0.790) argues against recombination events within this region. Such events would produce haplotypes that contained nucleotides that had not coevolved and therefore the nucleotide characters would not be consistent with a single phylogeny. The suppression of recombination within rDNA arrays in other mushroom species has been observed (Cassidy *et al.* 1984; Liou 2000), but recombination has been observed between arrays of the mushroom *L. bicolor* (Selosse *et al.* 1996; Martin *et al.* 1999). The observation of within-rDNA array homogenization combined with repression of homologous recombination between arrays is also reminiscent of recombination of rDNA in *Saccarhomyces cerevisiae* in which meiotic crossovers between sister chromatids are highly favored over nonsister exchanges (Petes and Pukkila 1995).

Our data support the notion that the nucleotide substitution rate is higher at the IGS2 region than at the IGS1 region in *S. commune*, which is in turn higher than that for the ITS spacer region. Data from the IGS regions suggest that the process of intrachromosomal homogenization of repeats on the same array occurs faster than the collective substitution rate of all the sites at IGS1. In turn, nucleotide substitution occurs faster than interchromosomal homogenization can make populations monomorphic for a single IGS1 haplotype. For the ITS region, the homogenization of arrays also happens faster than the substitution rate, the latter of which appears to progress almost as slowly as genetic drift and fixation of haplotypes within populations.

We thank Robert C. Ullrich for providing the *S. commune* strains from the collection of J. R. Raper. Thanks are also due to Cathie Aime, Omon Isikhuehmen, Karen Nakasone, David Porter, Martin Schnittler, and Sebastian Verduiin for providing additional material. John Wares, Stuart McDaniel, Cymon Cox, and Austen Ganley provided critical comments on earlier drafts. Two anonymous reviewers are also thanked for suggestions to improve the manuscript. Steve Leathers provided valuable laboratory assistance in sequencing efforts. This work was funded by a National Science Foundation Grant to R.V. (DEB-9408347) and a Sigma Xi GIAR to T.J.

LITERATURE CITED

Arnheim, N., 1983 Concerted evolution of multigene families, pp. 38–61 in *Evolution of Genes and Proteins*, edited by M. Nei and R. K. Koehn. Sinauer Associates, Boston.

Ásgeirsdóttir, S. A., F. H. J. Schuren and J. G. H. Wessels, 1994 Assignment of genes to pulse-field separated chromosomes of *Schizophyllum commune*. Mycol. Res. **98:** 689–693.

Carbone, I., J. B. Anderson and L. M. Kohn, 1999 Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. Evolution **53:** 11–21.

Cassidy, J. R., D. Moore, B. C. Lu and P. J. Pukkila, 1984 Unusual organization and lack of recombination in the ribosomal RNA genes of *Coprinus cinereus*. Curr. Genet. **8:** 607–613.

Cooke, W. B., 1961 The genus *Schizophyllum*. Mycologia **53:** 575–599.

Dons, J. J. M., and J. G. H. Wessels, 1980 Sequence organization of the nuclear DNA of *Schizophyllum commune*. Biochim. Biophys. Acta **607:** 385–396.

Dover, G. A., 1982 Molecular drive: a cohesive mode of species evolution. Nature **295:** 111–117.

Excoffier, L., P. E. Smouse and J. M. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:** 479–491.

Farris, J. S., 1989 The retention index and the rescaled consistency index. Cladistics **5:** 417–419.

Felsenstein, J., 1982 How can we infer geography and history from gene frequencies? J. Theor. Biol. **96:** 9–20.

Hirst, J. M., and G. W. Hurst, 1967 Long-distance spore transport. Symp. Soc. Gen. Microbiol. **17:** 307–344.

Hiscock, S. J., and U. Kües, 1999 Cellular and molecular mechanisms of sexual incompatibility in plants and fungi. Int. Rev. Cytol. **193:** 165–295.

Horak, E., 1983 Mycogeography in the South Pacific region: Agaricales, Boletales. Austr. J. Bot. Suppl. Ser. **10:** 1–41.

Ingold, C. T., 1971 *Fungal Spores: Their Liberation and Dispersal*. Oxford University Press, London.

James, T. Y., D. Porter, J. L. Hamrick and R. Vilgalys, 1999 Evidence for limited intercontinental gene flow in the cosmopolitan mushroom, *Schizophyllum commune*. Evolution **53:** 1665–1677.

Liou, S.-R., 2000 Evolutionary genetics of speciation in basidiomycetes: genetic studies of reproductive isolation in *Pleurotus djamor/calyptratus* complex. Ph.D. Thesis, Duke University, Durham, NC.

Martin, F., M.-A. Selosse and F. Le Tacon, 1999 The nuclear rDNA intergenic spacer of the ectomycorrhizal basidiomycete *Laccaria bicolor*: structural analysis and allelic polymorphism. Microbiology **145:** 1605–1611.

May, G., F. Shaw, H. Badrane and X. Vekemans, 1999 The signature of balancing selection: fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. USA **96:** 9172–9177.

Moncalvo, J.-M., F. M. Lutzoni, S. A. Rehner, J. Johnson and R. Vilgalys, 2000 Phylogenetic relationships of agaric fungi based on nuclear large subunit ribosomal DNA sequences. Syst. Biol. **49:** 278–305.

Nei, M., 1988 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neigel, J. E., 1997 A comparison of alternative strategies for estimating gene flow from genetic markers. Annu. Rev. Ecol. Syst. **28:** 105–128.

Nicholas, K. B., and H. B. Nicholas, Jr., 1997 GeneDoc: analysis and visualization of genetic variation. http://www.cris.com/~Ketchup/genedoc.shtml.

Pady, S. M., and L. Kapica, 1955 Fungi in air over the Atlantic Ocean. Mycologia **47:** 34–50.

Peterson, R. H., and S. C. McCleneghan, 1997 Reports on long-distance sexual compatibility in Agaricales. Nord. J. Bot. **17:** 419–432.

Petes, T. D., and P. J. Pukkila, 1995 Meiotic sister chromatid recombination. Adv. Genet. **33:** 41–62.

Posada, D., K. A. Crandall and A. R. Templeton, 2000 GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol. Ecol. **9:** 487–488.

Ramsfield, T. D., E. M. Becker, S. M. Rathlef, Y. Tang, T. C. Vrain *et al.*, 1996 Geographic variation of *Chondrostereum purpureum* detected by polymorphisms in the ribosomal DNA. Can. J. Bot. **74:** 1919–1929.

Raper, J. R., G. S. Krongelb and M. G. Baxter, 1958 The number and distribution of incompatibility factors in *Schizophyllum commune*. Am. Nat. **92:** 221–232.

Rihs, J. D., A. A. Padhye and C. B. Good, 1996 Brain abscess caused by *Schizophyllum commune*: an emerging basidiomycete pathogen. J. Clin. Microbiol. **34:** 1628–1632.

Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Schneider, S., J.-M. Kueffer, D. Roessli and L. Excoffier, 1997 ARLEQUIN: a software for population genetic data analysis. Ver 1.1. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Geneva.

Selker, E. U., E. Morzycka-Wróblewska, J. N. Stevens and R. L. Metzenberg, 1986 An upstream signal is required for in vitro transcription of *Neurospora* 5S RNA genes. Mol. Gen. Genet. **205:** 189–192.

Selosse, M.-A., G. Costa, C. Di Battista, F. L. Tacon and F. Martin, 1996 Meiotic segregation and recombination of the intergenic spacer of the ribosomal DNA in the ectomycorrhizal basidiomycete *Laccaria bicolor*. Curr. Genet. **30:** 332–337.

Singh, R. S., and L. R. Rhomberg, 1987 A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. Genetics **117:** 255–271.

Swofford, D. L., 1993 *PAUP: Phylogenetic Analysis Using Parsimony. 3.1.1*. User's manual. Smithsonian Institution, Washington, DC.

Swofford, D. L., 1998 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates, Sunderland, MA.

Templeton, A. R., 1998 Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. **7:** 381–397.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Templeton, A. R., E. Routman and C. A. Philips, 1995 Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. Genetics **140:** 767–782.

Ullrich, R. C., 1977 Natural distribution of incompatibility factors in basidiomycetous fungi. Mycologia **69:** 714–719.

Venkateswarlu, K., S.-W. Lee and R. N. Nazar, 1991 Conserved upstream elements in plant 5S ribosomal RNA encoding genes. Gene **105:** 249–253.

Vilgalys, R., and M. Hester, 1990 Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. J. Bacteriol. **172:** 4238–4246.

Vilgalys, R., and B. L. Sun, 1994 Ancient and recent patterns of geographic speciation in the oyster mushroom *Pleurotus* revealed by phylogenetic analysis of ribosomal DNA sequences. Proc. Natl. Acad. Sci. USA **91:** 4599–4603.

White, T. J., T. D. Bruns, S. B. Lee and J. W. Taylor, 1990 Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, pp. 315–321 in *PCR Protocols: A Guide to Methods and Applications*, edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White. Academic Press, New York.

Whitehouse, H. L. K., 1949 Multiple-allelomorph heterothallism in the fungi. New Phytol. **48:** 212–244.

Wright, S., 1939 The distribution of self-sterility alleles in populations. Genetics **24:** 538–552.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Zambino, P., J. V. Groth, L. Lukens, J. R. Garton and G. May, 1997 Variation at the b mating type locus of *Ustilago maydis*. Phytopathology **87:** 1233–1239.

Zolan, M. E., and P. J. Pukkila, 1986 Inheritance of DNA methylation in *Coprinus cinereus*. Mol. Cell. Biol. **6:** 195–200.

Communicating editor: S. Yokoyama