Evolution, 57(12), 2003, pp. 2703–2720

A MULTILOCUS GENEALOGICAL APPROACH TO PHYLOGENETIC SPECIES RECOGNITION IN THE MODEL EUKARYOTE *NEUROSPORA*

JEREMY R. DETTMAN,[1] DAVID J. JACOBSON,[2] AND JOHN W. TAYLOR[3]
Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3102
[1]*E-mail: jeremyd@uclink4.berkeley.edu*
[2]*E-mail: djjacob@stanford.edu*
[3]*E-mail: jtaylor@socrates.berkeley.edu*

Abstract.—To critically examine the relationship between species recognized by phylogenetic and reproductive compatibility criteria, we applied phylogenetic species recognition (PSR) to the fungus in which biological species recognition (BSR) has been most comprehensively applied, the well-studied genus *Neurospora*. Four independent anonymous nuclear loci were characterized and sequenced from 147 individuals that were representative of all described outbreeding species of *Neurospora*. We developed a consensus-tree approach that identified monophyletic genealogical groups that were concordantly supported by the majority of the loci, or were well supported by at least one locus but not contradicted by any other locus. We recognized a total of eight phylogenetic species, five of which corresponded with the five traditional biological species, and three of which were newly discovered. Not only were phylogenetic criteria superior to traditional reproductive compatibility criteria in revealing the full species diversity of *Neurospora*, but also significant phylogenetic subdivisions were detected within some species. Despite previous suggestions of hybridization between *N. crassa* and *N. intermedia* in nature, and the fact that several putative hybrid individuals were included in this study, no molecular evidence in support of recent interspecific gene flow or the existence of true hybrids was observed. The sequence data from the four loci were combined and used to clarify how the species discovered by PSR were related. Although species-level clades were strongly supported, the phylogenetic relationships among species remained difficult to resolve, perhaps due to conflicting signals resulting from differential lineage sorting.

Key words.—Biological species, genealogical concordance, hybridization, lineage sorting, phylogenetic species, species concepts, species recognition.

Received January 29, 2003. Accepted September 11, 2003.

Recognizing phylogenetic species using genealogical concordance of multiple independent loci (Avise and Ball 1990; Baum and Shaw 1995) has become feasible for most groups of organisms due to the increased ease of obtaining large amounts of nucleic acid sequence data. However, species delineation still is commonly based upon data from a single locus, in spite of clear evidence that a single gene genealogy does not necessarily represent the organismal phylogeny (Pamilo and Nei 1988; Takahata 1989; Avise and Wollenberg 1997; Rosenberg 2002). Furthermore, a single locus genealogy does not have the ability to detect the reticulation that is expected within a recombining species. The use of concordance of multiple gene genealogies to recognize species boundaries, hereafter referred to as phylogenetic species recognition (PSR), has been implemented in only a small number of empirical studies (e.g., Hilton and Hey 1997; Gleason et al. 1998; Kliman et al. 2000), the majority of which involve members of the fungal kingdom (e.g., Koufopanou et al. 1997; Geiser et al. 1998; Kasuga et al. 1999; O'Donnell et al. 2000a,b; Kroken and Taylor 2001; Cruse et al. 2002). A prevailing theme that emerges from these studies is that PSR recognizes additional genetically isolated species that had not been recognized previously, due to the lack of taxonomically informative morphological characters (phenotypic simplicity or plasticity) or incomplete reproductive isolation among species (cf. Taylor et al. 2000). Typically, a single morphological or biological species with a cosmopolitan distribution is found to be composed of multiple cryptic, phylogenetic species that often are geographically distinct. In addition, PSR is applicable to all organisms, including those that cannot be induced to mate in the laboratory, as is required for biological species recognition (BSR). For these reasons, PSR is becoming more popular, especially among mycologists, and is challenging BSR as the method of choice.

At this early stage in the application of genealogical concordance as a criterion to recognize phylogenetic species, we felt that it was important to critically examine the relationship between PSR and the current standard, BSR, by performing a broad and systematic comparison of the two methods in the model filamentous fungal genus *Neurospora* (Sordariales, Ascomycota). By independently implementing PSR and BSR on a common set of individuals, we were able to compare the empirical results of recognizing species under the frameworks of alternate species concepts. In this first of two reports, we describe the results of PSR using genealogical concordance of four independent nuclear loci. The results of BSR using mating tests and a thorough comparison to those of PSR are presented in a companion paper (Dettman et al. 2003).

Neurospora was chosen as the study organism because it is the most practical fungal system in which to compare the two species recognition methods. BSR has been more comprehensively applied to *Neurospora* than any other fungal group. Using well-established laboratory protocols (Perkins et al. 1976; Perkins and Turner 1988; Turner et al. 2001), reproductive success of crosses with species-specific tester strains has been used to assign most *Neurospora* individuals to one of five outbreeding biological species: *N. crassa*, *N. intermedia*, *N. sitophila*, *N. tetrasperma*, and *N. discreta*. In addition, *Neurospora* was chosen because relatively little research has addressed the evolutionary and natural history of this model organism, despite its extensive use for biochemical

genetics and molecular biology (Davis 2000; Davis and Perkins 2002), intensive, world-wide sampling of its natural populations (Turner et al. 2001; Jacobson et al. 2004), and the recently published genome sequence of *N. crassa* (Galagan et al. 2003).

The phylogenetic relationships among the five outbreeding species of *Neurospora* could not be inferred from morphological characters due to overlap among and inconsistency within species (Perkins et al. 1976; Perkins and Turner 1988; Turner et al. 2001). Genus-wide molecular phylogenies have shown that the five outbreeding species collectively form a monophyletic group (Pöggeler 1999; Dettman et al. 2001). Sequence data are available for six loci (*al-1, frq, gpd, mat a-1, mat A-1,* and ITS/5.8S rDNA; Randall and Metzenberg 1995; Skupski et al. 1997; Pöggeler 1999; Dettman et al. 2001), but the topologies of the trees produced from different loci support different relationships among *N. crassa*, *N. intermedia*, *N. tetrasperma*, and *N. sitophila*. Trees constructed from restriction fragment length polymorphism (RFLP) data from nuclear and mitochondrial DNA (Natvig et al. 1987; Taylor and Natvig 1989; Skupski et al. 1997) agreed with trees from only one of the six sequenced loci, *mat a-1*. The phylogenetic relationships among these four species still are equivocal, mainly because the internal branches (internodes) of the molecular phylogeny are, on average, only a quarter of the length of branches leading to species. These facts suggest that *N. crassa*, *N. intermedia*, *N. tetrasperma*, and *N. sitophila* arose during a rapid radiation, allowing for differential and/or incomplete sorting of ancestral polymorphism (Natvig and May 1996; Skupski et al. 1997). When all data are considered, the phylogeny remains unclear and the only consistent pattern is that *N. discreta* diverged first among the five species.

We focused here on two species of *Neurospora*, *N. crassa* and *N. intermedia*, because they represented a challenge to both BSR and PSR. With respect to BSR, *N. crassa* and *N. intermedia* are assumed to be closely related sibling species between which reproductive isolation may not be complete. Laboratory crosses of *N. crassa* × *N. intermedia* may produce a small but significant number of viable hybrid progeny. In addition, several individuals collected from nature have been described as possible hybrids of *N. crassa* and *N. intermedia* (Turner et al. 2001) because they mate moderately well with both the *N. crassa* and *N. intermedia* tester strains, in a fashion similar to that of true *N. crassa*/*N. intermedia* hybrids created in the laboratory. With respect to PSR, all previous studies using molecular data, albeit with relatively few individuals per species, have failed to resolve *N. crassa* and *N. intermedia* into two reciprocally monophyletic species. Restriction fragment length polymorphism data from random nuclear DNA (Natvig et al. 1987; Skupski et al. 1997) suggested that *N. intermedia* is a paraphyletic group, whereas RFLP data from mitochondrial DNA (Taylor and Natvig 1989) suggested both *N. crassa* and *N. intermedia* are polyphyletic. Sequence data from two genes (*al-1* and *frq*) further suggested polyphyly of *N. intermedia* (Skupski et al. 1997). If these *Neurospora* species did arise during a rapid radiation, they pose an even greater challenge to species delineation using molecular approaches.

To recognize phylogenetic species of *Neurospora*, we compared the genealogical patterns of four independent nuclear loci sampled from a large collection of individuals. We developed a consensus-tree approach that identified monophyletic genealogical groups that were concordantly supported by the majority of the loci, or were well supported by at least one locus but not contradicted by any other locus. To address the possibility of hybridization between *N. crassa* and *N. intermedia* in nature, putative *N. crassa*/*N. intermedia* hybrids, and *N. crassa* and *N. intermedia* individuals collected sympatrically, were included in these analyses. Sequence data from the four loci were combined and analyzed together to clarify the phylogenetic relationships among the *Neurospora* species.

Materials and Methods

Taxon Sampling

We chose 147 individuals (Appendix) to represent the five outbreeding, conidiating (mitotically sporulating) species of *Neurospora*. The majority of individuals were sampled from *N. crassa* (48) and *N. intermedia* (71), particularly from regions where *N. crassa* and *N. intermedia* are commonly sympatric, such as India and the Caribbean Basin. Sixty-two *N. intermedia* individuals were of the standard orange ecotype, and nine were of the yellow ecotype (Turner 1987). Nine putative *N. crassa*/*N. intermedia* hybrid individuals were selected, as were seven *N. sitophila*, four *N. tetrasperma*, and eight *N. discreta* individuals. Single individuals from four of the self-fertile (homothallic) *Neurospora* species (*N. dodgei*, *N. galapagosensis*, *N. africana*, and *N. lineolata*; FGSC 1692, 1739, 1740, and 1910, respectively) were included to root phylogenetic trees. Most strains were obtained directly from the Perkins collection (D. D. Perkins, Department of Biological Sciences, Stanford University, Stanford, CA). Additional strains were obtained from the personal collection of DJJ, or from the Fungal Genetics Stock Center (FGSC; Department of Microbiology, University of Kansas Medical Center, Kansas City, KS), which now maintains and curates the entire Perkins collection. To ensure that individual haploid genotypes were being characterized, a single germinated conidium was isolated from each received culture and the resulting subcultures were given unique identification numbers (D1–D147). This collection of single-conidium strains was redeposited in the FGSC and assigned new FGSC numbers (8761–8907).

Marker Development

Various microsatellite repeat motifs were used to search unassembled cosmid sequences from *N. crassa* linkage groups II and V (Shulte et al. 2002), which were available from the Munich Information Centre for Protein Sequences *Neurospora crassa* Database (http://www.mips.biochem.mpg.de/proj/neurospora/). Primer sets for twenty putative loci were designed to amplify a 400 to 700-bp fragment that included a microsatellite and flanking regions. Three loci from which a single band was amplified in a set of reference individuals from all five outbreeding species were chosen for further characterization. A fourth primer set was designed to amplify a microsatellite locus on linkage group IV from a

TABLE 1. Primer sequences, annealing temperatures, and genomic locations of loci. MNCDB, Munich Information Centre for Protein Sequences *Neurospora crassa* Database. WICGR, Whitehead Institute-Center for Genome Research.

Locus	Primers (5′-3′)	Annealing temperature (°C)	MNCDB contig	WICGR contig/ supercontig	Chromosomal location
TMI[1]	Forward = TMI-UPS: CACCCTCAGTATCTTCAACA, Reverse = TMI-3: TGTGAAGGTTGAGAGTATGG	54	9A18	3.51/3	Linkage group VR, between *rca-1* and *his-1*
DMG	Forward = DMG-5: GACGTCGCGCTATGCTCTGC, Reverse = DMG-3: TTTGGTCGGAATGGTCGGTG	59	cosmid G6G8[2]	3.379/29	Linkage group IVR, near *pyr-1*
TML	Forward = TML-UPS-B: GTCGGACACGAAGTGGACAA, Reverse = TML-3: AATCCCGCTTAGCAAAGGTG	58	9A71	3.555/61	Linkage group VL (predicted by MNCDB)
QMA	Forward = QMA-UPS: TCTTGATGGGAATTTATGTGA, Reverse = QMA-DWN-C: CCTAGGTTCCTATCTAGCCAG, Reverse = QMA-DWN-B: ATATGTGCCTAAAAGCAATCA	52	B15I20	3.307/20	Linkage group IIR, near *aro-3*

[1] Preliminary annotation of the *N. crassa* genome suggested the TMI locus may encode a hypothetical protein (NCU01303.1), but the existence of this hypothetical protein has not been verified.
[2] From Bean et al. 2001; GenBank AF309689.

cosmid clone described by Bean et al. (2001). Loci were sequenced from reference individuals and iterative rounds of primer design were performed until the four loci could be amplified and sequenced from most individuals in our sample. The Whitehead Institute Center for Genome Research *Neurospora crassa* Database (http://www-genome.wi.mit.edu/annotation/fungi/neurospora/) became available after characterization of the four loci and was used to verify the chromosomal locations of loci (Table 1).

DNA Extraction, Amplification, and Sequencing

Strains were grown in minimal medium broth (Vogel 1964) with 1% sucrose for 2–3 days at 34°C in an orbital shaker. Mycelial tissue was harvested by vacuum filtration, rinsed with distilled water, submerged in liquid nitrogen, lyophilized, and ground into a powder. Dry tissue (20–30 mg) was incubated at 65°C for 45 min in 500 μl of lysis buffer with final concentrations of 50 mM Tris-HCl, 50 mM EDTA, and 3% SDS. After a wash with 500 μl chloroform:isoamyl alcohol (24:1), samples were centrifuged (12,000 × *g*, 5 min) to remove cellular debris. The aqueous phase was collected and genomic DNA was extracted using the DNeasy Tissue Kit (Qiagen Inc., Valencia, CA; crude lysates protocol). Loci were PCR-amplified from genomic DNA with the following reaction conditions: 200 μM dNTPs, 0.4 μM each primer (Operon Technologies Inc, Alameda, CA), 1.0 unit of AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA), 1X PCR buffer (supplied with enzyme). The thermal cycler protocol was as follows: initial denaturation at 94°C for 2 min; 34 cycles of 94°C for 1 min, locus-specific annealing temperature (Table 1) for 1 min, 72°C for 1 min; 7 min extension at 72°C; maintenance at 4°C. Amplification products were purified using the QIAquick PCR Purification Kit (Qiagen) and nucleotide sequences were determined using BigDye Terminator Cycle Sequencing Kits and a 377 DNA Sequencer (Applied Biosystems). Cloning of PCR products was not necessary because our *Neurospora* strains were haploid and possessed only one allele per individual. Sequence data were collected from both strands and were examined using Sequencing Analysis and Sequence Navigator (ver. 3.4 and ver. 1.0.1, respectively; Applied Biosystems). Nucleotide sequences of the four loci from a representative individual from each phylogenetic species and intraspecific subgroup (and each homothallic species, if sequenced) have been deposited in GenBank under accessions AY225899–AY225949.

Phylogenetic Analyses

DNA sequences were aligned preliminarily using ClustalX (ver. 1.8; Thompson et al. 1997; gap opening = 10, gap extension = 0.2, transition weight = 0.5) then edited visually. Regions of sequence with ambiguous alignment were excluded from all analyses (Table 2) and gaps were treated as missing data. For parsimony analyses, insertion/deletion gaps (indels) that were consistently and unambiguously alignable across all taxa were treated as single evolutionary events by recoding a single site within the indel as a multistate character. Owing to the possibility of microsatellite allele length homoplasy, microsatellite repeats themselves were excluded from all phylogenetic analyses and will be presented in a future report. After microsatellites were removed, flanking sequences were concatenated to create the final alignments (Table 2). The full alignment containing all four loci has been deposited in TREEBASE under accessions S950 and M1574. Kimura two-parameter genetic distances (d = number of nucleotide differences per site between two sequences; Kimura 1980) were calculated using MEGA (ver. 2.1; Kumar et al. 2001).

To avoid confinement at local optima in maximum parsimony (MP) searches, we performed 100 replicates of ran-

dom stepwise-addition heuristic searches (nearest-neighbor interchange [NNI] branch-swapping; maximum of 10,000 trees retained [maxtrees]) using PAUP (ver. 4.0b8a; Swofford 2001). The shortest trees from 100 replicates were used as starting trees for more intensive MP heuristic searches (tree bisection-reconnection [TBR] branch swapping; maxtrees = 5000). For weighted MP analysis of the combined dataset, characters from the TMI, DMG, TML, and QMA loci were weighted as 0.46, 1.00, 0.34, and 0.22, respectively (i.e., inversely proportional to the total number of phylogenetically informative sites per locus). When multiple MP trees were produced, the tree chosen for display in a figure was the one determined to be most likely using substitution models suggested by ModelTest (ver. 3.06; Posada and Crandall 1998). Maximum parsimony bootstrapping was performed with heuristic searches (100 replicates for individual loci, 300 replicates for combined analysis; simple stepwise addition; NNI; maxtrees = 5000). Partition homogeneity tests (PHT), or incongruence length difference tests (Farris et al. 1995), were performed to detect significant conflict between phylogenetic signals of loci. To avoid detecting incongruence that is expected within lineages, PHT were restricted to datasets containing only 26 individuals that represented the main lineages (strains D1, D9, D11, D14, D16, D17, D35, D37, D39, D53, D55, D58, D70, D75, D78, D84, D89, D103, D106, D121, D129, D133, D140, D143, D145, and D146). Partition homogeneity tests used informative characters only and random stepwise-addition MP heuristic searches with 100 or 1000 replicates (TBR; maxtrees = 500). A significance threshold of 0.05 causes PHT to be too conservative (Cunningham 1997; Darlu and Lecointre 2002), so the null hypothesis of congruence was rejected only if $P < 0.001$.

Owing to computational constraints, maximum likelihood analyses could not be performed on our large datasets. Instead, Bayesian analyses were performed using MrBayes (ver. 3.0; Huelsenbeck and Ronquist 2001). Each run consisted of three to four incrementally heated Markov chains run simultaneously, with heating values set to default (0.2). Default uniform priors were used for all model parameters (six substitution rates, four base frequencies, proportion of invariable sites, and alpha value of gamma distribution). Markov chains were initiated from a random tree and were run for 500,000 or one million generations (single-locus or combined datasets, respectively), and samples were taken every 100th generation. Log-likelihood values were plotted against generation number and ''burn-in'' (i.e., lack of improvement of log-likelihood values) was evaluated visually. All samples taken prior to burn-in were discarded and the remaining samples were used to determine posterior probability distributions. To ensure analyses were not trapped in local optima, each run was performed independently at least twice. Log-likelihood values and consensus trees from stationary samples from each replicate run were compared to verify that they converged on congruent phylogenetic trees. For simplicity, only the replicate runs with the highest mean log-likelihood values were reported here.

RESULTS

Polymorphism Summary

Approximately 273,000 nucleotides of new sequence data are reported here. A total of 593 sequences were obtained, with an average of 460 nucleotides per sequence. Summaries of the alignments for the four loci (Table 1) characterized in this study are shown in Table 2. The loci differed strikingly in the amount and form of variation. For instance, QMA, the most variable locus, and DMG, the least variable locus, differed tenfold in the average genetic distance between any two individuals ($d = 0.084$ and 0.008, respectively) and fourfold in the number of phylogenetically informative characters (202 and 45, respectively). As predicted by the putative nonfunctional status of these loci, they contained numerous indels that ranged from one to 12 nucleotides in length. For TMI, only one of 98 (1.0%) phylogenetically informative characters was an indel, but for DMG, 16 of 45 (35.6%) phylogenetically informative characters were indels, the highest proportion for any locus.

Not surprisingly, the four noncoding loci sequenced in this study were on average 74.7% more variable than the six previously characterized functional genes. For TMI, DMG, TML, and QMA combined, the mean genetic distance between pairs of individuals in interspecific comparisons of the five biological species was 0.058, compared to 0.033 for *al-1, frq, gpd, mat a-1, mat A-1,* and ITS/5.8S rDNA combined (data from Randall and Metzenberg 1995; Skupski et al. 1997; Pöggeler 1999; Dettman et al. 2001).

Analyses of Single-Locus Datasets

To take advantage of numerous informative indels present in the data, maximum parsimony (MP) was used to infer genealogies from the four single-locus alignments (TMI: tree length = 174 steps, consistency index [CI] = 0.845; DMG: tree length = 102 steps, CI = 0.941; TML: tree length = 363 steps, CI = 0.713; QMA: tree length = 518 steps, CI = 0.633). The full 147-taxon trees were too large to display here, so only schematics of the four single-locus tree are presented (Fig. 1). The full trees are in an Electronic Appendix, currently available from the *Evolution* Editorial office at evolution@asu.edu. As displayed in Figure 1, there were several well-supported clades common to the single-locus trees. Based upon these trees, and the analyses described below, several groups of individuals were named for the sake of discussion: *N. crassa* subgroups NcA, NcB, and NcC; *N. intermedia* subgroups NiA and NiB; and Phylogenetic Species 1, 2, and 3 (PS1, PS2, and PS3).

Bayesian methods (Huelsenbeck and Ronquist 2001) were used to accommodate more complex evolutionary models and thereby account for differences in base frequencies, substitution types, and substitution rates among sites. Relationships among the main clades were topologically identical in MP and Bayesian consensus trees; therefore, Bayesian trees are not shown. Instead, the Bayesian posterior probabilities (PP) of branches (or clades) are included on the trees in Figure 1.

Sequence data were obtained for homothallic *Neurospora* species, which represent lineages outside the monophyletic clade of outbreeding species. The TMI locus was sequenced from *N. galapagosensis*, *N. africana*, *N. dodgei*, and *N. lineolata*, and the DMG locus was sequenced from the latter three species. When these homothallic taxa were assigned to the outgroup instead of *N. discreta*, the root of the outbreeding species tree was placed along the branch that led to *N. discreta*

TABLE 2. Summary of the five DNA sequence alignments.

Locus	TMI	DMG	TML	QMA	Combined
Number of strains	147	147	147	145[1]	147
Length of final alignment	446	465	681	549[2]	2141
Excluded positions due to ambiguous alignment	none	428–443	158–163 and 435–452	none	874–889, 1069–1074, and 1346–1363
Position of omitted microsatellites	after base 382[3]	after base 279	after base 649	after bases 449 and 467	after bases 382, 725, 1560, 2041, and 2059
Average number of included nucleotides per sequence	445	378	538	482	1836
Total number of phylogenetically informative characters (substitutions, indels)	98 (97, 1)	45 (29, 16)	132 (98, 34)	202 (181, 21)	477 (405, 72)
Average genetic distance, *d* (SE)	0.0371 (0.0053)	0.0081 (0.0020)	0.0294 (0.0039)	0.0839 (0.0078)	0.0393 (0.0021)

[1] The QMA locus could not be amplified from strains D104 and D138.
[2] *Neurospora discreta* sequences after base 434 were unalignable with sequences from other taxa, so bases 435–549 were scored as missing for *N. discreta*.
[3] For example: bases 1–382 were upstream and bases 383–446 were downstream from the microsatellite prior to its removal from the alignment.

(trees not shown). This rooting was consistent with phylogenies based on sequence data from the previously characterized functional loci (*gpd, mat a-1, mat A-1,* and ITS/5.8S rDNA; Pöggeler 1999; Dettman et al. 2001); thus, all current evidence supports the use of *N. discreta* as the outgroup for the other outbreeding species.

To determine whether phylogenetic signals were incongruent between loci, partition homogeneity tests (PHT) were performed for all six possible pairings of loci. To focus on relationships among species and subgroups, and to avoid detecting incongruence that is expected within lineages due to recombination, datasets were pruned to contain only 26 individuals that were representative of the main lineages (six from *N. crassa* [two from each of NcA, NcB, and NcC], eight from *N. intermedia* [four from NiA, two from NiB, and two basal lineages], and two from each of PS1, PS2, PS3, *N. sitophila*, *N. tetrasperma*, and *N. discreta*; see Materials and Methods for strain numbers). Significant conflict was detected for only two pairs of loci: QMA-TMI and QMA-TML ($P < 0.001$). Note that these two comparisons contained QMA, the most variable locus, and the next two most variable loci. QMA was the anomalous locus because phylogenetic signals of TMI, DMG, and TML were congruent in all paired combinations. As displayed in Figure 1, the main difference between QMA and the other loci was the placement of PS2 in relation to the other species.

Analyses of Combined Dataset

The allelic genealogies differed among the individual loci (Fig. 1 and Electronic Appendix), and none necessarily represented the organismal phylogeny, particularly if a rapid radiation had occurred recently. For inferring the organismal phylogeny, we feel that all data should be considered together, whether the loci are fully congruent or not. A tree produced from a MP heuristic search using the combined alignment of 2141 characters from all four loci is shown in Figure 2. In this analysis, the QMA locus had the greatest influence on the tree search because it contributed 42.4% of the total phylogenetically informative characters. To allow each locus to contribute equally, a weighted MP heuristic search (tree length = 506.20, CI = 0.640, tree not shown) was performed with characters weighted inversely proportional to the total number of phylogenetically informative sites contributed by the locus from which they came. The relative branching order of ingroup species was the same in the MP and weighted MP trees, and the only relevant difference in branch support was increased support for monophyly of *N. intermedia* in weighted MP analysis (MP bootstrap proportion [MPBP] of 94%).

The relationships among species and subgroups were topologically identical in the tree produced from Bayesian analysis of the combined dataset (mean log-likelihood value = −10832.23, tree not shown) and both MP trees. Bayesian PP values of main branches are displayed on the MP tree in Figure 2. Three more replicate Bayesian analyses were run, all of which had marginally lower mean log-likelihood values (−10837.23, −10845.58, and −10850.83) than the first run. Nonetheless, all four replicate runs converged on similar consensus trees with the same relative branching order of the ingroup species.

Phylogenetic Species Recognition by Genealogical Concordance

After speciation, the neutral genetic polymorphism shared between the newly diverged species is gradually lost through genetic drift. The species pass through stages of polyphyly and paraphyly on the way to eventual reciprocal monophyly (Avise and Ball 1990). Owing to the stochasticity of this process, each locus will lose shared ancestral polymorphism and become fixed for derived polymorphism at a different rate. Thus, for closely related species, exclusive monophyly of all sampled loci is an unreasonably strict criterion for phylogenetic species recognition (Hudson and Coyne 2002; Rosenberg 2003).

Grouping of individuals.—A clade was recognized as an independent evolutionary lineage if it satisfied either of two

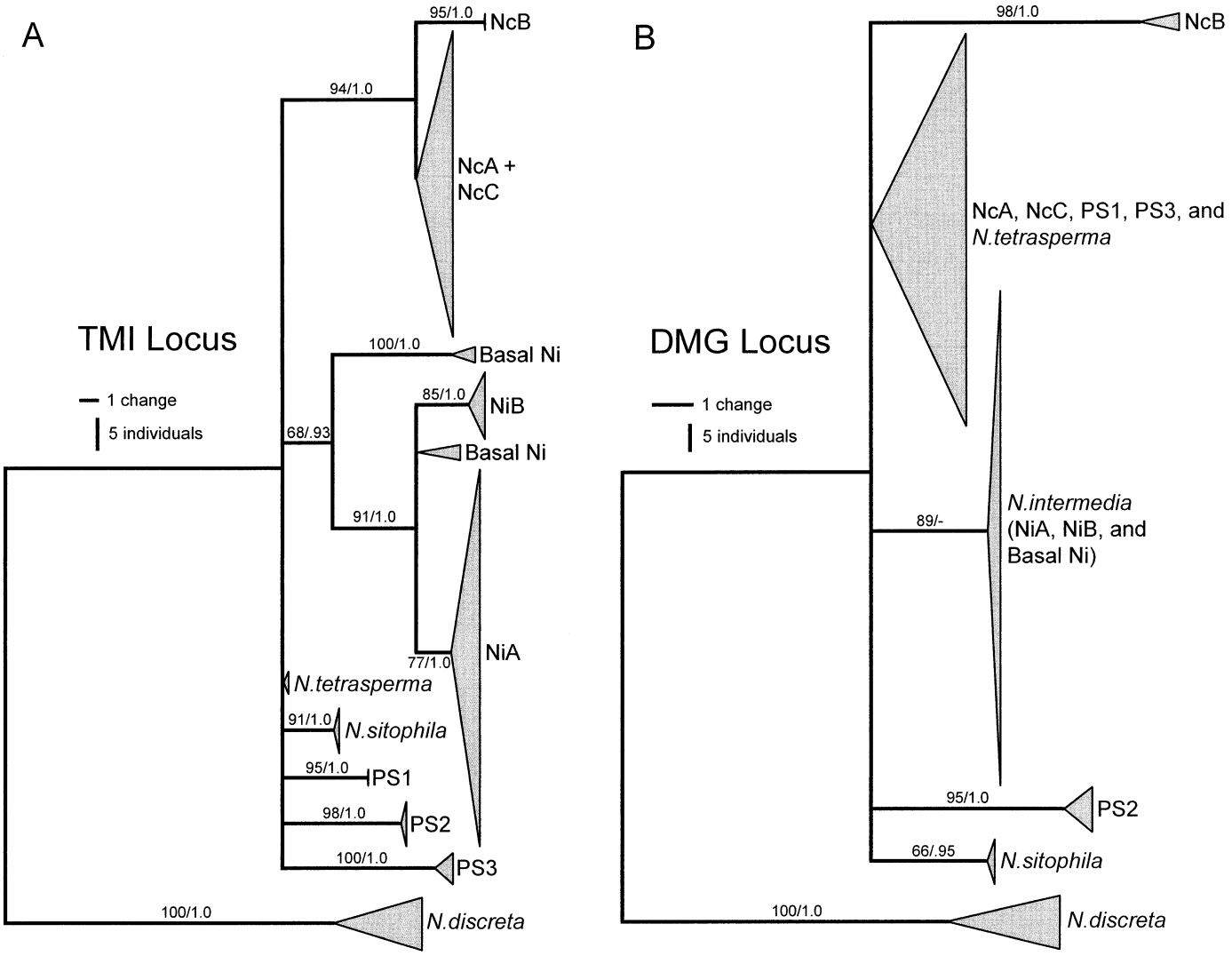


FIG. 1. Schematic representations of the maximum parsimony phylograms produced from each of the four single-locus datasets. The well-supported groups of individuals that appear in common in multiple single-locus trees are indicated by triangles, with height proportional to number of individuals and width proportional to the mean number of changes from the node (see scales). For simplicity, only the major branches of the trees are displayed, with poorly supported branches reduced to polytomies (i.e., triangles with zero branch lengths). Branch support values are indicated by numbers near branches (maximum parsimony bootstrap proportions/Bayesian posterior probabilities). The full phylograms are accessible as supplemental material in an Electronic Appendix, currently available from the *Evolution* Editorial office at evolution@asu.edu.

criteria: (1) Genealogical concordance: the clade was present in the majority (3/4) of the single-locus genealogies. To identify such clades, a majority-rule consensus tree was produced from the four single-locus trees (see Electronic Appendix). This criterion revealed the genealogical patterns shared among loci, regardless of levels of support. (2) Genealogical nondiscordance: the clade was well supported in at least one single-locus genealogy, as judged by both MP bootstrap proportions (Hillis and Bull 1993) and Bayesian posterior probabilities (Rannala and Yang 1996; Larget and Simon 1999), and was not contradicted in any other single-locus genealogy at the same level of support. To identify such clades, a tree possessing only branches that received a MPBP $\geq$ 70% and a Bayesian PP $\geq$ 0.95 was chosen to represent each of the four loci, then a semistrict consensus tree (combinable component; Bremer 1990) was produced from these four trees. This criterion prohibited poorly supported nonmonophyly at one locus from undermining well-supported monophyly at another locus.

In Figure 2, bold branches indicate the clades that satisfied either of these grouping criteria and therefore were identified as independent evolutionary lineages.

Ranking of groups.—When deciding which independent evolutionary lineages represented phylogenetic species, characteristics of lineages in combined data analyses were also considered. Two ranking criteria were applied: (1) Genetic differentiation: to prevent minor tip clades from being recognized, phylogenetic species had to be relatively distinct and well differentiated from other species. (2) Exhaustive subdivision: all individuals had to be placed within a phylogenetic species. The purpose of this study was to partition a set of individuals into phylogenetic species to compare

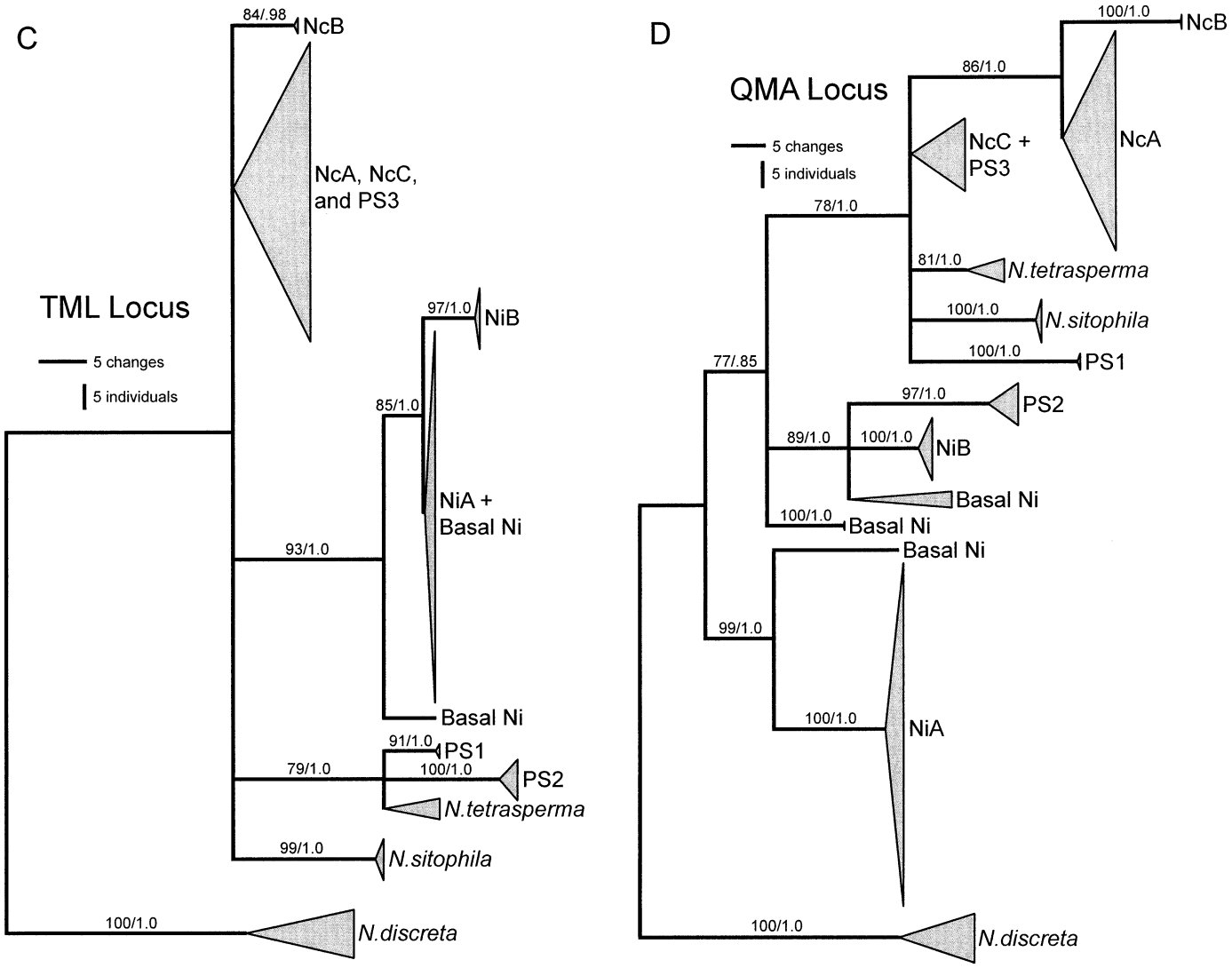
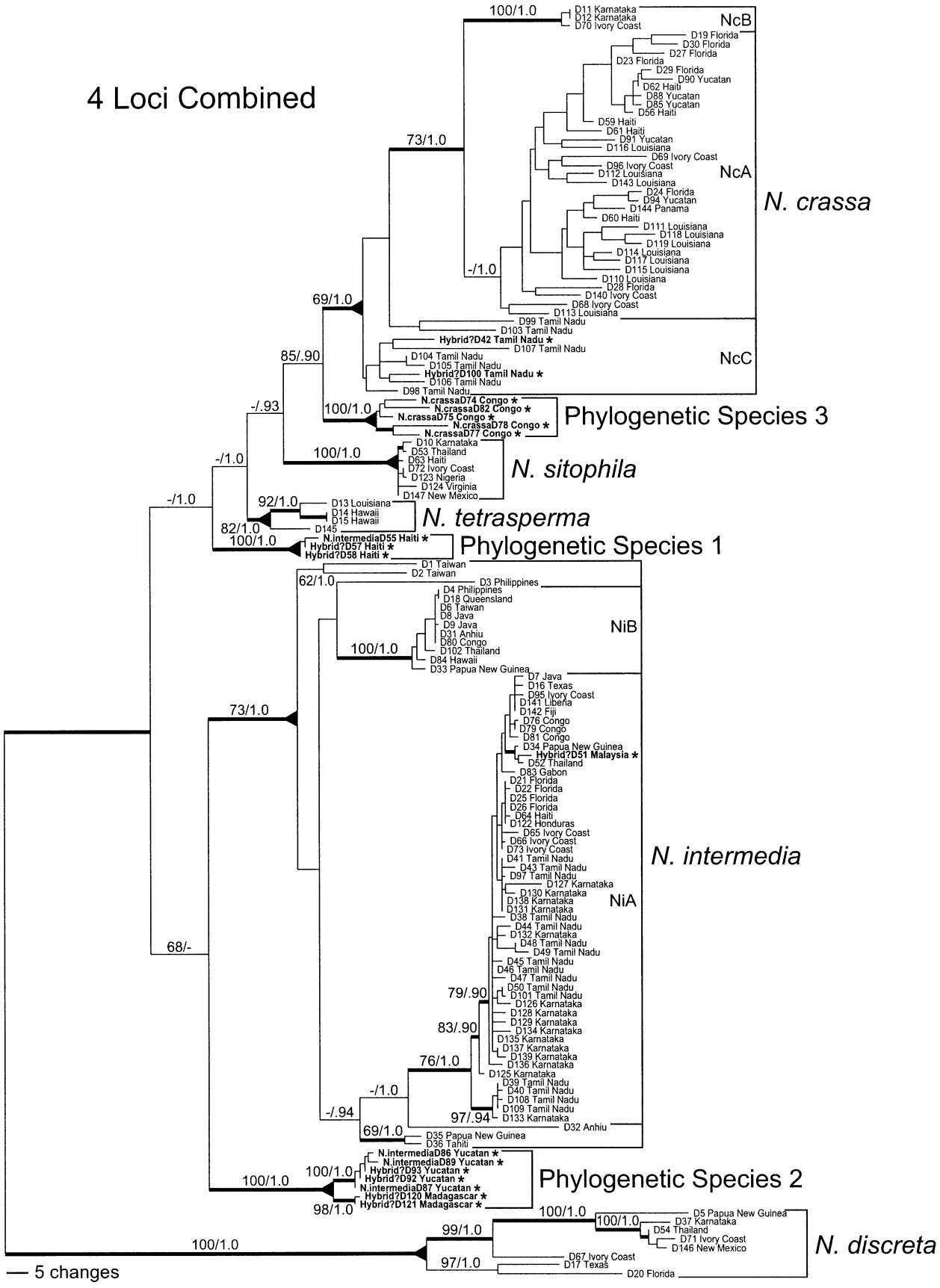


FIG. 1. Continued.

these groups with those delineated through biological species recognition (Dettman et al. 2003), so no individuals were to be left unclassified. First, all terminal independent evolutionary lineages (bold branches in Fig. 2) were identified. Then, if an individual was not included within one of these clades, we traced down the nodes of the tree from that individual, collapsing clades not subtended by bold branches. We continued this until all individuals were included in a clade subtended by a bold branch, and recognized such clades as phylogenetic species. A consequence of these criteria was increased inclusiveness of recognized species. For example, the latter criterion would cause an independent lineage and its nonmonophyletic sister group to be recognized together as a single species.

Based on the grouping and ranking criteria, within what was originally regarded as five biological species, we recognized eight phylogenetic species as described below:

Neurospora crassa.—The phylogenetic species *N. crassa* was composed of 45 strains, and monophyly of the entire *N. crassa* clade was well supported in the analyses of the combined loci (69%:1.0 = MPBP:Bayesian PP) and TMI (94%: 1.0), and was not contradicted significantly by any other locus. *Neurospora crassa* contained a significant amount of internal phylogenetic structure with three subgroups: NcA, NcB, and NcC. The majority of *N. crassa* strains were included within the NcA subgroup, which had intermediate support in combined analysis. This subgroup consisted of all *N. crassa* strains from the Caribbean Basin and all but one from Africa. Also included in NcA was strain D143, the wild-type laboratory strain of *N. crassa* (74-OR23-1A) that has been used extensively in genetic and biochemical studies. A second small, distinct subgroup, NcB, was well supported in all single-locus (84–100%:0.98–1.0) and combined analyses (100%:1.0). Nine strains from Tamil Nadu, India, formed NcC, a third, paraphyletic subgroup that was basal to other *N. crassa* subgroups. Although NcC strains typically were clustered together, their monophyly was not well supported in any analyses and their placement in relation to other groups


4 Loci Combined
N. crassa
NcB
NcA
NcC
Phylogenetic Species 3
N. sitophila
N. tetrasperma
Phylogenetic Species 1
N. intermedia
NiB
NiA
Phylogenetic Species 2
N. discreta
5 changes

varied in the single-locus trees (e.g., mixed within NcA and/or PS3; Fig. 1). Two strains within the NcC subgroup (D42 and D100) had been described as putative *N. crassa*/*N. intermedia* hybrids (Turner et al. 2001).

Although the NcB subgroup was an independent lineage as judged by our grouping criteria, it was not recognized as a species because doing so would have resulted in NcA and NcC not being assigned to any species. Because NcA and NcC were not identified as independent lineages, we recognized NcA + NcB + NcC as a single phylogenetic species with three subgroups.

Neurospora intermedia.—The phylogenetic species *N. intermedia* comprised 68 strains, and monophyly of the entire *N. intermedia* clade was well supported in combined analysis (73%:1.0; 94% MPBP in weighted MP) and received intermediate to significant support from three of the four individual loci (TMI, DMG, and TML). Like *N. crassa*, *N. intermedia* had internal phylogenetic structure with two main subgroups identified. The large subgroup NiA was well supported in TMI, QMA, and combined-analysis trees (76–100%:1.0) and was composed of 52 strains, one of which (D51) had been considered a putative *N. crassa*/*N. intermedia* hybrid. Another group of ten strains formed a subgroup, NiB, which had significant support in combined-analysis (100%:1.0) and all single-locus trees (85–100%:1.0) except the low-resolution DMG tree. Nine strains described as the ''yellow ecotype'' of *N. intermedia* (Turner 1987) were included in this study, and all nine fell within NiB. A tenth strain, D8, consistently fell within the NiB clade but had not been identified as a yellow ecotype strain because it lacked the characteristic morphological attribute of large, yellow conidia. Under closer examination, this strain had an extremely slow growth rate and produced few to no conidia (probably due to infection by a virus; Tuveson and Peterson 1972), which precluded proper identification by morphology.

Six strains (D1, D2, D3, D32, D35, and D36) were not included in NiA or NiB and formed long-branched, basal lineages in TMI, QMA, and combined-analysis trees. In the DMG and TML trees, the placement of these lineages within *N. intermedia* as a whole was well supported. However, the relationships among these lineages were not consistent among the single-locus trees or the different methods of combined analyses, and they did not form a monophyletic group in any analyses. Although the NiA and NiB subgroups were both identified as independent evolutionary lineages, they were not recognized as two separate species because doing so would have left the remaining basal lineages unclassified. These basal lineages could not be recognized as a single species or multiple species because neither alternative met our grouping criteria. Therefore, we recognized *N. intermedia* as a single, large phylogenetic species (NiA + NiB + basal lineages). *Neurospora intermedia* was the only phylogenetic species whose monophyly was contradicted by a well-supported branch: NiB was more closely related to PS2 than to NiA at the anomalous QMA locus.

Phylogenetic Species 1 (PS1).—A group of three strains from Haiti formed PS1, which was well supported in combined analyses (100%:1.0) and all single-locus trees (91–100%:1.0) except the low-resolution DMG locus. Two PS1 strains had been considered putative hybrids, and the other had been described as a possible *N. intermedia* strain (Turner et al. 2001).

Phylogenetic Species 2 (PS2).—PS2 was composed of seven strains and was well supported in all single-locus (95–100%:1.0) and combined analyses (100%:1.0). Within PS2 itself, there were two closely related, well-supported subgroups that reflected the geographic sources of the strains. The first subgroup was composed of five strains collected from the Yucatan Peninsula, which previously were identified as two *N. intermedia*, one possible *N. intermedia*, and two putative hybrids. The second subgroup consisted of two putative hybrids from Madagascar. Because the two subgroups within PS2 lacked significant genetic differentiation and were composed of a few individuals from geographically disjunct sites, we collapsed them into a single species to reflect that their reciprocal monophyly may be compromised with the addition of further accessions from other localities.

Phylogenetic Species 3 (PS3).—Five strains collected from the Republic of the Congo formed PS3, which was monophyletic and well supported by TMI and combined analysis (both 100%:1.0). In the DMG, TML, and QMA trees, PS3 strains were intermingled with *N. crassa* strains, mostly from the NcC subgroup (Tamil Nadu, India). However, the placement of PS3 strains within *N. crassa* was not well supported by any analyses.

Neurospora tetrasperma, N. sitophila, and N. discreta.—A well-supported monophyletic group of four *N. tetrasperma* strains was present in the combined-analysis (82%:1.0) and QMA trees (81%:1.0). The seven *N. sitophila* strains formed a well-supported monophyletic group in all analyses (66–100%:0.95–1.0), as did the eight *N. discreta* strains (100%:1.0). Although the sample size for *N. discreta* was small compared to other species, it was the most genetically variable and contained four well-supported subdivisions. With further study, it is likely that multiple phylogenetic species will be discovered within this diverse group of *N. discreta* strains.

←

FIG. 2. Maximum parsimony phylogram produced from the TMI, DMG, TML, and QMA loci combined (tree length = 1369, consistency index = 0.606.). Labels to the right of the phylogram indicate groups identified by phylogenetic analyses. Bold branches were concordantly supported by the majority of the loci, or were well supported by at least one locus but not contradicted by any other locus. Triangles at nodes indicate that all taxa united by (or distal to) it belong to the same phylogenetic species (see text for details). Taxon labels indicate strain number and geographic source. If a strain was originally identified by traditional mating tests to a species that did not match the phylogenetic species identification, the original species name is listed before the strain number, and is followed by an asterisk, all in bold face. If the original species identification matched the phylogenetic species identification, no name appears before the strain number. Branch support values (maximum parsimony bootstrap proportions/Bayesian posterior probabilities; MPBP/PP) in combined analyses are displayed for major branches only. A dash indicates the support for the branch was <50% MPBP or <0.50 PP.

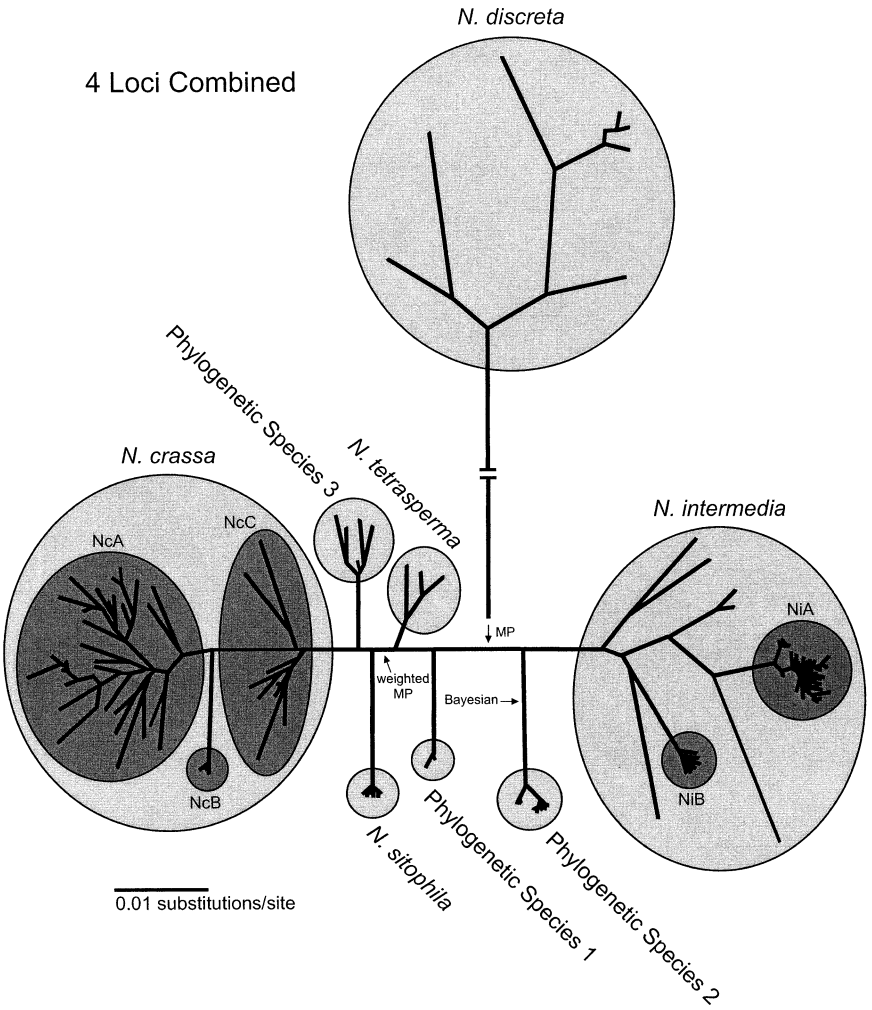


FIG. 3. An unrooted tree summarizing the relationships among phylogenetic species based on combined data. All branch lengths are proportional except the branch leading to *Neurospora discreta*, which is longer than indicated. Phylogenetic species are enclosed within lightly shaded circles, and subgroups are enclosed in darkly shaded circles. Each terminal branch represents an individual. All methods of phylogenetic reconstruction (maximum parsimony [MP], weighted MP, and Bayesian) produced trees with the same relative branching order of the ingroup taxa as shown here, but which differed in the placement of the ingroup root, as indicated by arrows.

Relationships among Phylogenetic Species

The eight well-supported phylogenetic species (and subgroups therein) described above were present in multiple single-locus trees (Fig. 1). Overall, there were no conflicts among the four loci in terms of which individuals belonged in which phylogenetic species. However, the phylogenetic relationships among species were difficult to resolve. The internal branches that united multiple species, or species internodes, typically received nonsignificant support (Fig. 1 and Electronic Appendix) and no sister-group relationship was well supported by more than one locus.

When the species phylogeny was estimated from analyses of the four loci combined, some species internodes were well supported by at least one method of branch support estimation (Fig. 2). The relationship among species that received the most support was the placement of PS3 sister to *N. crassa*. In combined analysis, the PS3-*N. crassa* branch was supported by a MPBP of 85% and a PP of 0.90. To assess support for the branch that separated PS2 and *N. intermedia* from the other ingroup species, we compensated for the uncertainty of the ingroup root (see below) by repeating bootstrap analyses with *N. discreta* omitted. In these analyses, the PS2-*N. intermedia* sister-group relationship received significant support (MPBP of 73%, PP of 1.0). Regardless of the phylogenetic inference method (MP, weighted MP, or Bayesian) used to construct the combined-analysis tree, the relative branching order of the seven ingroup phylogenetic species was the same (Fig. 3).

Overall, the species internodes were relatively short and poorly supported compared to branches that led to species (Figs. 2 and 3). To explore the possibility that incongruence of phylogenetic signals between loci was responsible for the short internal branches with poor support, we repeated the analyses with the three congruent loci combined (TMI, DMG, and TML). The exclusion of QMA, the most divergent locus, did not significantly increase the bootstrap support of any of the species internodes. Therefore, the lack of resolution among phylogenetic species was not due simply to conflict

among the genealogies, but rather was a characteristic of the genealogies themselves (i.e., short species internodes).

The placement of the ingroup root, or the branch of the ingroup tree that was bisected by the branch that led to the *N. discreta* outgroup, differed among single-locus trees and among combined-analysis trees produced by different methods. To reflect this uncertainty, we display an unrooted tree in Figure 3 and indicate the rooting possibilities suggested by the different methods of phylogenetic reconstruction.

Discussion

Phylogenetic Species Recognition

We subscribe to the evolutionary species concept sensu Simpson (1951) in which a species is viewed as ''a single lineage of ancestral descendent populations of organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate'' (Wiley 1978, p. 18). The idea that species are lineages (de Queiroz 1998) is essential to the evolutionary species concept, and to most other species concepts that have been proposed (Mayden 1997). In general, the several methods of species recognition all attempt to identify evolutionary lineages, but use different criteria to do so (Avise and Wollenberg 1997; Mayden 1997).

We recognized phylogenetic species using genealogical concordance and nondiscordance of four nuclear loci that represented unlinked segments of the *Neurospora* genome. Within a single interbreeding species, the mixing effects of recombination would cause unlinked loci to have incongruent genealogies (Tajima 1983; Pamilo and Nei 1988; Takahata 1989; Rosenberg 2002), but between genetically isolated species, the extinction of ancestral alleles by genetic drift would lead to the congruence of genealogies. Therefore, the transition between deep genealogical concordance and shallow genealogical discordance can be used to recognize phylogenetic species (Avise and Ball 1990; Baum and Shaw 1995; Avise and Wollenberg 1997; Taylor et al. 2000). The loss of ancestral polymorphism due to genetic drift causes newly diverged species to pass through stages of polyphyly and paraphyly on the way to eventual reciprocal monophyly (Avise and Ball 1990). Most *Neurospora* species passed our tests of consistently supported monophyly or well-supported and uncontradicted monophyly, which indicated that the barriers to genetic exchange have existed for a long period relative to the population sizes of the species (Hudson and Coyne 2002; Rosenberg 2003).

Our application of PSR was conservative because not every independent evolutionary lineage (bold branches in Fig. 2) was designated a phylogenetic species (triangles at nodes in Fig. 2). For both *N. crassa* and *N. intermedia*, multiple distinct subgroups and/or lineages were recognized together as a single phylogenetic species. We understand that other authors might have considered each *N. crassa* subgroup to be a separate species, despite poor support for NcA and the paraphyly of NcC (which would be a ''metaspecies''; Donoghue 1985). Similarly, others might have considered NiA and NiB to be two separate species, and left the basal *N. intermedia* lineages unclassified, or described each basal lineage as its own species. Regardless of the subjectivity in taxonomic practice, the important implications of this study are that phylogenetically defined groups of individuals exist, and the relationships among them can and should be considered when these individuals are included in future studies of the evolutionary biology of *Neurospora*.

Biological Species Recognition and the Lack of True Hybrids

With traditional BSR, newly collected *Neurospora* individuals are assigned to biological species based on the reproductive success of crosses to species tester strains using refined laboratory protocols (Perkins et al. 1976; Perkins and Turner 1988; Turner et al. 2001). We recognized eight phylogenetic species, although the 147 individuals of *Neurospora* included in this study had been placed into only five biological species by traditional BSR. Each of the five previously described biological species corresponded to a well-supported phylogenetic species, and three additional distinct phylogenetic species (PS1, PS2, and PS3) were discovered. This outcome supports the conclusion that using genealogical concordance as the criterion for species recognition generally results in the identification of more species than using reproductive isolation as the criterion.

For 129 (87.8%) of the 147 individuals, the original species designations determined by traditional BSR were equivalent to the phylogenetic species designations determined in this study. Excluding the putative hybrids, only nine individuals were incorrectly identified by traditional BSR (Figure 2 and Appendix). One PS1 and three PS2 individuals had been placed in *N. intermedia*, and five PS3 individuals had been placed in *N. crassa* by traditional BSR.

Turner et al. (2001) described several individuals collected from nature that could not be assigned unequivocally to a species by traditional BSR. These individuals displayed partial fertility with tester strains from both *N. crassa* and *N. intermedia*, suggesting that they may represent natural hybrids (Turner et al. 2001). When multiple unlinked loci are sequenced, true hybrid individuals would be expected to possess *N. crassa*-like alleles at some loci and *N. intermedia*-like alleles at others. The phylogenetic placement of each of the nine putative hybrids was consistent across all four loci, thus providing no evidence in support of their hybrid status. Individuals that mate well with multiple biological species have been reported in other groups of organisms (e.g., Gleason et al. 1998; Aanen et al. 2000). Typically, molecular evidence from multiple loci has placed the problematic individuals into one species or the other, comparable to our results with these nine putative hybrids.

Two of the nine putative hybrids included in this study were assigned by PSR to the NcC subgroup of *N. crassa*. Another putative hybrid fell consistently within the NiA subgroup of *N. intermedia*. These three individuals were no more genetically distinct from their respective *N. crassa* and *N. intermedia* species testers than any other individuals, so their misidentification by BSR was presumably due to strain-specific sexual reproductive irregularities. The remaining six putative hybrids were assigned to either PS1 or PS2 (Figure 2 and Appendix). The fact that appropriate PS1 or PS2 testers strains did not exist at the time of attempted identification

clearly explains the equivocal species assignment of these individuals by traditional BSR. Herein lies a distinct advantage of PSR over BSR: PSR is less dependent upon prior knowledge of the existence of a species. Traditional BSR was quite successful at placing unknown individuals into a existing species framework, but was limited at detecting rare or under-sampled species. Further discussion of the relationship between PSR and BSR appears in Dettman et al. (2003).

Geographic Differentiation and Intraspecific Trends

Neurospora crassa was the species that displayed the most geographic differentiation. Within the NcA subgroup, phylogenetic relationships among the geographically distributed alleles presented no clear evidence for differentiation among populations throughout the Caribbean Basin or Africa. On the other hand, it appeared that little genetic exchange has occurred recently between the NcC (India) and NcA subgroups. Consistent with this observation, between-subgroup crosses (NcC × NcA) were generally less fertile than within-subgroup crosses (NcC × NcC, or NcA × NcA; Dettman et al. 2003). The evidence for limited genetic exchange and partial reproductive isolation both suggested that the NcC and NcA subgroups were incipient species.

The *N. intermedia* subgroup NiB corresponded to the yellow ecotype of *N. intermedia*. In addition to possessing large, saffron yellow conidia, this ecotype is collected almost exclusively from nonburned substrate, whereas the more common orange ecotype is typically found on recently burned plant material (Turner 1987; Turner et al. 2001). The sequence data presented here suggested limited genetic exchange between the NiA and NiB subgroups in the recent past. The yellow and orange ecotypes have diverged morphologically, ecologically, and phylogenetically, but significant reproductive isolation between ecotypes has not been detected (Turner 1987). Although the yellow ecotype did show reduced fertility when crossed with the orange ecotype, Turner (1987) concluded that the two ecotypes were conspecific because some, but not all, crosses between yellow and orange ecotype strains were highly fertile.

Descendents of strains D1 and D2 were chosen as *N. intermedia* testers for traditional BSR due to their high fertility with the majority of other *N. intermedia* strains (Shew 1978; Turner et al. 2001). Interestingly, these two individuals formed a divergent basal *N. intermedia* lineage that was phylogenetically differentiated from the main bulk of *N. intermedia* sampled worldwide (Fig. 2). Evidently, the ability to mate successfully is an ancestral character that has been retained by these groups despite significant phylogenetic divergence.

Individuals placed within the same newly discovered phylogenetic species typically were sampled from the same geographic location (e.g., PS1 from Haiti, PS3 from the Congo). The genus *Neurospora* appears to have both geographically widespread species and narrowly distributed endemics. Curiously, PS2 displayed ecological divergence from other species because all seven PS2 individuals were sampled from nonburned substrate (soil or dry leaf material). As with the yellow ecotype of *N. intermedia*, it is possible that differential substrate preference has contributed to divergence of PS2 from other lineages.

The five PS3 individuals were recognized as a phylogenetic species distinct from *N. crassa*. Fixed genetic differences between these two species were detected at only one locus, but fixed genetic differences between PS3 and NcA existed at three of the four loci. However, all PS3 individuals were highly fertile with *N. crassa* testers (Turner et al. 2001) and several other *N. crassa* individuals (Dettman et al. 2003). The genetic differentiation between Ivory Coast NcA strains and PS3 strains, which are all from the Congo, suggested the presence of some geographic barrier between these two sites. However, no evidence for such a geographic barrier was found in *N. intermedia*, in which little differentiation between strains from the Congo and Ivory Coast was detected. If PS3 and NcA individuals can mate in nature, factors other than reproductive or geographic isolation are preventing genetic exchange between them.

Phylogeny of Outbreeding Neurospora Species

The genealogies constructed from the four nuclear loci were not fully congruent (Fig. 1), and significant conflict was detected between some loci. This observation emphasizes the fact that a genealogy reflects the branching order of the allelic lineages of that locus, not necessarily the branching order of the populations or species from which it was sampled. The importance of assessing the genealogical patterns of several independent loci prior to drawing conclusions about a species phylogeny was underscored by our study and by other empirical studies that have also clearly demonstrated that different nuclear loci can have different genealogies (e.g., Hilton and Hey 1997; Geiser et al. 1998; Hare and Avise 1998, Piercey-Normore et al. 1998; Carbone et al. 1999; Kasuga et al. 1999; Aanen et al. 2000; Kliman et al. 2000; O'Donnell et al. 2000a,b; Xu et al. 2000; Kroken and Taylor 2001; Hare et al. 2002). In general, increased numbers of sampled loci are associated with increased accuracy and confidence of phylogenetic inference (Pamilo and Nei 1988; Takahata 1989; Rosenberg 2002).

Although the four nonfunctional and presumably neutral loci characterized in this study contributed more polymorphic sites than previously characterized functional loci (*al-1, frq, gpd, mat a-1, mat A-1,* and ITS/5.8S rDNA; Randall and Metzenberg 1995; Skupski et al. 1997; Pöggeler 1999; Dettman et al. 2001), the *Neurospora* phylogeny still is not fully resolved. The short, internal branches that united multiple species typically received greater support from Bayesian PP than from MP bootstrapping (Fig. 2). If Bayesian support values are more accurate than bootstrap values, as concluded by Alfaro et al. (2003) and Wilcox et al. (2002), then more trust may be placed in the internal branches of the phylogeny. In contrast to all previous molecular studies of *Neurospora*, the sequence data presented here were able to resolve *N. crassa* and *N. intermedia* into two monophyletic species. In summary, the tree displayed in Figure 3 is our best hypothesis of the phylogenetic relationships among species.

These eight phylogenetic species of *Neurospora* have been evolving independently long enough to allow a significant amount of fixed nucleotide differences to accumulate among

species, as indicated by the well-supported long branches that led to species. The general lack of shared alleles among lineages suggested that gene flow among species has not occurred in the recent past. When compared to the long branches leading to species, it appears that the time between speciation events was short relative to the time since the last speciation event, assuming a constant mutation rate.

Neurospora as a Model Organism for the Study of Evolutionary Biology

A reliable and detailed phylogenetic framework is a prerequisite for addressing other questions regarding evolution and speciation in *Neurospora*. As a result of this study and the companion study also published in this issue (Dettman et al. 2003), researchers can place individuals of *Neurospora* into any of several species, and clades therein, which now are well characterized with respect to hierarchical phylogenetic relationships and reproductive compatibility. We hope that this information, and the attributes of *Neurospora* described in the introduction, will further the establishment of *Neurospora* as a model system for the study of evolution, speciation, and population biology.

Acknowledgments

We thank E. Turner and D. Baum for critical review of the manuscript, and D. Perkins for his inspiration and dedication to the study of natural populations of *Neurospora*. Funding for this research was provided by the National Science Foundation grant DEB-9981987 to JWT. This paper is based in part on the Ph.D. dissertation of JRD at the University of California, Berkeley.

Literature Cited

Aanen, D. K., T. W. Kuyper, T. H. M. Mes, and R. F. Hoekstra. 2000. The evolution of reproductive isolation in the ectomycorrhizal *Hebeloma crustuliniforme* aggregate (Basidiomycetes) in northwestern Europe: a phylogenetic approach. Evolution 54: 1192–1206.

Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20:255–266.

Avise, J. C., and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Pp. 45–67 *in* D. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

Avise, J. C., and K. Wollenberg. 1997. Phylogenetics and the origin of species. Proc. Natl. Acad. Sci. USA 94:7748–7755.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 *in* P. C. Hoch and A. G. Stephenson, eds. Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis, MO.

Bean, L. E., W. H. Dvorachek, E. L. Braun, A. Errett, G. S. Saenz, M. D. Giles, M. Werner-Washburne, M. A. Nelson, and D. O. Natvig. 2001. Analysis of the *pdx-1* (*snz-1*/*sno-1*) region of the *Neurospora crassa* genome: correlation of pyridoxine-requiring phenotypes with mutations in two structural genes. Genetics 157: 1067–1075.

Bremer, K. 1990. Combinable component consensus. Cladistics 6: 369–378.

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.

Cruse, M., R. Telerant, T. Gallagher, T. Lee, and J. W. Taylor. 2002. Cryptic species in *Stachybotrys chartarum*. Mycologia 94: 814–822.

Cunningham, C. 1997. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 14:733–740.

Darlu, P., and G. Lecointre. 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19:432–437.

Davis, R. H. 2000. *Neurospora*: contributions of a model organism. Oxford Univ. Press, New York.

Davis, R. H., and D. D. Perkins. 2002. *Neurospora*: a model of model microbes. Nat. Rev. Genet. 3:7–13.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Pp. 57–75 *in* D. J. Howard and S. H. Berlocher, eds. Endless forms: species and speciation. Oxford Univ. Press, Oxford, U.K.

Dettman, J. R., F. M. Harbinski, and J. W. Taylor. 2001. Ascospore morphology is a poor predictor of the phylogenetic relationships of *Neurospora* and *Gelasinospora*. Fungal Genet. Biol. 34:49–61.

Dettman, J. R., D. J. Jacobson, E. Turner, A. Pringle, and J. W. Taylor. 2003. Reproductive isolation and phylogenetic divergence in *Neurospora*: comparing methods of species recognition in a model eukaryote. Evolution 57:2721–2741.

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist 88:172–181.

Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics 10:315–319.

Galagan, J., and 76 others. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859–868.

Geiser, D. M., J. I. Pitt, and J. W. Taylor. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. Proc. Natl. Acad. Sci. USA 95:388–393.

Gleason, J. M., E. C. Griffith, and J. R. Powell. 1998. A molecular phylogeny of the *Drosophila willistoni* group: Conflicts between species concepts? Evolution 52:1093–1103.

Hare, M. P., and J. C. Avise. 1998. Population structure in the American oyster as inferred by nuclear gene genealogies. Mol. Biol. Evol. 15:119–128.

Hare, M. P., F. Cipriano, and S. R. Palumbi. 2002. Genetic evidence on the demography of speciation in allopatric dolphin species. Evolution 56:804–816.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

Hilton, H., and J. Hey. 1997. A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. Genet. Res. 70:185–194.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Jacobson, D. J., A. J. Powell, J. R. Dettman, G. S. Saenz, M. M. Barton, M. D. Hiltz, W. H. Dvorachek, N. L. Glass, J. W. Taylor, and D. O. Natvig. 2004. *Neurospora* in temperate forests of western North America. Mycologia 96:66–74.

Kasuga, T., J. W. Taylor, and T. J. White. 1999. Phylogenetic relationships of varieties and geographical groups of the human pathogenic fungus *Histoplasma capsulatum* Darling. J. Clin. Microbiol. 37:653–663.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics 156:1913–1931.

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.

Kroken, S., and J. W. Taylor. 2001. A gene genealogical approach to recognize phylogenetic species boundaries in the lichenized fungus *Letharia*. Mycologia 93:38–53.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245.
Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.
Mayden, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381–424 *in* M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds. Species: the units of biodiversity. Chapman and Hall, London.
Natvig, D. O., and G. May. 1996. Fungal evolution and speciation. J. Genet. 75:441–452.
Natvig, D. O., D. A. Jackson, and J. W. Taylor. 1987. Random-fragment hybridization analysis of evolution in the genus *Neurospora*: the status of four-spored strains. Evolution 41: 1003–1021.
O'Donnell, K., H. C. Kistler, B. K. Tacke, and H. H. Casper. 2000a. Gene genealogies reveal global phylogeographic structure and reproductive isolation among lineages of *Fusarium graminearum*, the fungus causing wheat scab. Proc. Natl. Acad. Sci. USA 97:7905–7910.
O'Donnell, K., H. I. Nirenberg, T. Aoki, and E. Cigelnik. 2000b. A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. Mycoscience 41:61–78.
Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.
Perkins, D. D., and B. C. Turner. 1988. *Neurospora* from natural populations: toward the population biology of a haploid eukaryote. Exp. Mycol. 12:91–131.
Perkins, D. D., B. C. Turner, and E. G. Barry. 1976. Strains of *Neurospora* collected from nature. Evolution 30:281–313.
Piercey-Normore, M. D., K. N. Egger, and J. A. Berube. 1998. Molecular phylogeny and evolutionary divergence of North American biological species of *Armillaria*. Mol. Phylogenet. Evol. 10:49–66.
Pöggeler, S. 1999. Phylogenetic relationships between mating-type sequences from homothallic and heterothallic ascomycetes. Curr. Genet. 36:222–231.
Posada, D., and K. A. Crandall. 1998. ModelTest: testing the model of DNA substitution. Bioinformatics 14:817–818.
Randall, T. A., and R. L. Metzenberg. 1995. Species-specific and mating type-specific DNA regions adjacent to mating type idiomorphs in the genus *Neurospora*. Genetics 141:119–136.
Rannala, B., and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43:304–311.
Rosenberg, N. A. 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61:225–247.
———. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution 57:1465–1477.
Schulte, U., I. Becker, H. W. Mewes, and G. Mannhaupt. 2002. Large scale analysis of sequences from *Neurospora crassa*. J. Biotech. 94:3–13.
Shew, H. W. 1978. New mutants and wild-type standard reference strains of *Neurospora intermedia*. Neurospora Newsl. 25:25.
Simpson, G. G. 1951. The species concept. Evolution 5:285–298.
Skupski, M. P., D. A. Jackson, and D. O. Natvig. 1997. Phylogenetic analysis of heterothallic *Neurospora* species. Fungal Genet. Biol. 21:153–162.
Swofford, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.
Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460.
Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122:957–966.
Taylor, J. W., and D. O. Natvig. 1989. Mitochondrial DNA and evolution of heterothallic and pseudohomothallic *Neurospora* species. Mycol. Res. 93:257–272.
Taylor, J. W., D. J. Jacobson, S. Kroken, T. Kasuga, D. M. Geiser, D. S. Hibbett, and M. C. Fisher. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21–32.
Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
Turner, B. C. 1987. Two ecotypes of *Neurospora intermedia*. Mycologia 79:425–432.
Turner, B. C., D. D. Perkins, and A. Fairfield. 2001. *Neurospora* from natural populations: a global study. Fungal Genet. Biol. 32:67–92.
Tuveson, R. W., and J. F. Peterson. 1972. Virus-like particles in certain slow-growing strains of *N. crassa*. Virology 47:527–531.
Vogel, H. J. 1964. Distribution of lysine pathways among fungi: evolutionary implications. Am. Nat. 98:435–466.
Wilcox, T. P., D. J. Zwickl, T. A. Heath, and D. M. Hillis. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25:361–371.
Wiley, E. O. 1978. The evolutionary species concept reconsidered. Syst. Zool. 27:17–26.
Xu, J. P., R. Vilgalys, and T. G. Mitchell. 2000. Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. Mol. Ecol. 9:1471–1481.

Corresponding Editor: D. Baum

APPENDIX

Strains of *Neurospora* subjected to phylogenetic species recognition.

Strain numbers[1]							Geographic location		
D	FGSC	Old FGSC	Other	Mating type	Original species[2]	Phylogenetic species[3]	Region[4]	Collection site	Substrate[5]
1	8761	1766[6]	13	*A*	*N. intermedia*	*N. intermedia*	East Asia	Taipei, Taiwan	grass
2	8762	1767	17	*a*	*N. intermedia*	*N. intermedia*	East Asia	Taipei, Taiwan	grass
3	8763	1763	25	*a*	*N. intermedia*	*N. intermedia*	East Asia	Manila, Philippines	straw
4	8764	5641	27	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Manila, Philippines	maize cob, unburnt
5	8765		47[6]	*a*	*N. discreta*	*N. discreta*	East Asia	Tiaba, Papua New Guinea	grass
6	8766	1791	61	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Sing-in, Taiwan	cane pressing, unburnt
7	8767	1792	142	*A*	*N. intermedia*	*N. intermedia* (NiA)	East Asia	Bogor, Java	grass
8	8768	2215	147	*a*	*N. intermedia*	*N. intermedia* (NiB)	East Asia	Bogor, Java	ontjom, unburnt
9	8769	5645	176	*a*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Tjipanas, Java	maize cob, unburnt
10	8770	2491	333	*A*	*N. sitophila*	*N. sitophila*	India	Golur, Karnataka	Euphorb
11	8771		334	*a*	*N. crassa*	*N. crassa* (NcB)	India	Golur, Karnataka	Euphorb
12	8772		335	*a*	*N. crassa*	*N. crassa* (NcB)	India	Golur, Karnataka	Euphorb
13	8773	2503[6]	505	*A*	*N. tetrasperma*	*N. tetrasperma*	Carib. Basin	Welsh, Louisiana	grass
14	8774	2510[6]	556	*A*	*N. tetrasperma*	*N. tetrasperma*	East Asia	Hanalei, Hawaii	grass
15	8775	2509[6]	581	*a*	*N. tetrasperma*	*N. tetrasperma*	East Asia	Lihue, Hawaii	grass
16	8776	3213[6]	831	*a*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Fred, Texas	shrub
17	8777	3228[6]	851	*A*	*N. discreta*	*N. discreta*	Carib. Basin	Kirbyville, Texas	grass
18	8778	5647	898	*a*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	8 Mile Plain, Queensland, Australia	filter mud, unburnt
19	8779	3971	1405	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
20	8780		1406	*A*	*N. discreta*	*N. discreta*	Carib. Basin	Homestead, Flordia	grass
21	8781	6603	1407	*A*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Homestead, Flordia	grass
22	8782		1408	*a*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Homestead, Flordia	grass
23	8783	3970	1409	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
24	8784		1410	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
25	8785		1413	*a*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Homestead, Flordia	grass
26	8786		1415	*A*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Homestead, Flordia	grass
27	8787		1417	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
28	8788		1460	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
29	8789		1465	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Flordia	grass
30	8790		1470	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Homestead, Florida	grass
31	8791	3980	1515	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Hefei, China	maize cob, unburnt
32	8792	3989	1526	*A*	*N. intermedia*	*N. intermedia*	East Asia	Hefei, China	maize cob, unburnt
33	8793	5339	1737	*a*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Goroka, Papua New Guinea	maize cob, unburnt
34	8794		1794	*a*	*N. intermedia*	*N. intermedia* (NiA)	East Asia	Baiyer River, Papua New Guinea	grass
35	8795	4876	1878	*a*	*N. intermedia*	*N. intermedia*	East Asia	Tiaba, Papua New Guinea	grass
36	8796	6582	2362	*A*	*N. intermedia*	*N. intermedia*	East Asia	Arue, Tahiti (Moorea)	grass
37	8797	6789	2511	*a*	*N. discreta*	*N. discreta*	India	Bandipur, Karnataka	soil isolation, unburnt
38	8798	4718	2533	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
39	8799		2536	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
40	8800		2538	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
41	8801		2542	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
42	8802	8198	2543	*a*	putative hybrid	*N. crassa* (NcC)	India	Madaurai, Tamil Nadu	sugarcane
43	8803		2544	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Konappatti, Tamil Nadu	sugarcane
44	8804	5344	2546	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Konappatti, Tamil Nadu	sugarcane
45	8805		2550	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Konappatti, Tamil Nadu	sugarcane
46	8806		2551	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Konappatti, Tamil Nadu	sugarcane

APPENDIX. Continued.

Strain numbers[1]				Mating type	Original species[2]	Phylogenetic species[3]	Geographic location		Substrate[5]
D	FGSC	Old FGSC	Other				Region[4]	Collection site	
47	8807	5345	2552	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Rameshwaram, Tamil Nadu	sugarcane
48	8808		2554	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Rameshwaram, Tamil Nadu	sugarcane
49	8809	5346	2555	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Rameshwaram, Tamil Nadu	sugarcane
50	8810		2567	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Mallilinatham, Tamil Nadu	sugarcane
51	8811	8199	2632	*A*	putative hybrid	*N. intermedia* (NiA)	East Asia	Georgetown, Malaya (Penang)	grass
52	8812	5369	2938	*A*	*N. intermedia*	*N. intermedia* (NiA)	East Asia	Ban Khao Yai, Thailand	shrub
53	8813	6489	2998	*A*	*N. sitophila*	*N. sitophila*	East Asia	Khao Yai, Thailand	shrub
54	8814	6792	3016	*a*	*N. discreta*	*N. discreta*	East Asia	Kang Koi, Thailand	grass
55	8815		3423	*a*	*N. intermedia*?	PS1	Carib. Basin	Carrefour Dufort, Haiti	grass
56	8816	4710	3424	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Carrefour Dufort, Haiti	grass
57	8817	8200	3425	*A*	putative hybrid	PS1	Carib. Basin	Carrefour Dufort, Haiti	grass
58	8818	8225	3426	*A*	putative hybrid	PS1	Carib. Basin	Carrefour Dufort, Haiti	grass
59	8819		3427	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Carrefour Dufort, Haiti	sugarcane
60	8820	4712	3433	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Leogane, Haiti	sugarcane
61	8821	4713	3437	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Merger, Haiti	shrub
62	8822	4824	3491	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Carrefour Mme. Gras, Haiti	shrub
63	8823	4791	3494	*A*	*N. sitophila*	*N. sitophila*	Carib. Basin	Carrefour Mme. Gras, Haiti	shrub
64	8824	6251	3495	*A*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Carrefour Mme. Gras, Haiti	shrub
65	8825	6254	3540	*A*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Yopougon, Ivory Coast	grass
66	8826	6255	3543	*a*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Yopougon, Ivory Coast	grass
67	8827		3660	*a*	*N. discreta*-like	*N. discreta*	Africa	Fougbesso, Ivory Coast	shrub
68	8828	4825	3681	*A*	*N. crassa*	*N. crassa* (NcA)	Africa	Tiassale, Ivory Coast	grass
69	8829	4826	3684	*a*	*N. crassa*	*N. crassa* (NcA)	Africa	Tiassale, Ivory Coast	grass
70	8830	4830	3726	*A*	*N. crassa*	*N. crassa* (NcB)	Africa	Golikro, Ivory Coast	grass
71	8831		3728	*A*	*N. discreta*?	*N. discreta*	Africa	Golikro, Ivory Coast	grass
72	8832	6262	3758	*A*	*N. sitophila*	*N. sitophila*	Africa	Adiopodoume, Ivory Coast	grass
73	8833	6263	3770	*a*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Adiopodoume, Ivory Coast	grass
74	8834		3812	*a*	*N. crassa*	PS3	Africa	Makaba, Congo	grass
75	8835	4821	3816	*a*	*N. crassa*	PS3	Africa	Makaba, Congo	grass
76	8836	6271	3824	*a*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Loubomo, Congo	grass
77	8837	4820	3826	*a*	*N. crassa*	PS3	Africa	Loubomo, Congo	grass
78	8838	4822	3838	*A*	*N. crassa*	PS3	Africa	Madingo, Congo	grass
79	8839	4274	3839	*a*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Madingo, Congo	grass
80	8840		3843	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	Africa	Madingo, Congo	grass
81	8841	6276	3852	*a*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Bouanza, Congo	grass
82	8842	4819	3853	*a*	*N. crassa*	PS3	Africa	Bouanza, Congo	grass
83	8843	6286	3932	*A*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Makokou, Gabon	grass
84	8844		4095	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Laie, Oahu, Hawaii	shrub
85	8845		4130	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Kabah, Yucatan, Mexico	soil isolation, unburnt
86	8846		4146	*a*	*N. intermedia*	PS2	Carib. Basin	Merida, Yucatan, Mexico	soil isolation, unburnt

APPENDIX. Continued.

Strain numbers[1]							Geographic location		
D	FGSC	Old FGSC	Other	Mating type	Original species[2]	Phylogenetic species[3]	Region[4]	Collection site	Substrate[5]
87	8847		4148	*A*	*N. intermedia*	PS2	Carib. Basin	Merida, Yucatan, Mexico	soil isolation, unburnt
88	8848		4150	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Sayil, Yucatan, Mexico	soil isolation, unburnt
89	8849		4151	*a*	*N. intermedia*?	PS2	Carib. Basin	Kabah, Yucatan, Mexico	soil isolation, unburnt
90	8850		4154	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Uxmal, Yucatan, Mexico	soil isolation, unburnt
91	8851		4155	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Uman, Yucatan, Mexico	soil isolation, unburnt
92	8852	8201	4157	*a*	putative hybrid	PS2	Carib. Basin	Merida, Yucatan, Mexico	soil isolation, unburnt
93	8853	8202	4158	*a*	putative hybrid	PS2	Carib. Basin	Merida, Yucatan, Mexico	soil isolation, unburnt
94	8854		4160	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Chichen Itza, Yucatan, Mexico	soil isolation, unburnt
95	8855	6857	4277	*A*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Adiopodoume, Ivory Coast	grass
96	8856		4305	*A*	*N. crassa*	*N. crassa* (NcA)	Africa	Adiopodoume, Ivory Coast	Euphorb
97	8857		4331	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Konappatti, Tamil Nadu	sugarcane
98	8858		4333	*A*	*N. crassa*	*N. crassa* (NcC)	India	Mallilinatham, Tamil Nadu	sugarcane
99	8859		4334	*A*	*N. crassa*	*N. crassa* (NcC)	India	Mallilinatham, Tamil Nadu	sugarcane
100	8860	8203	4335	*a*	putative hybrid	*N. crassa* (NcC)	India	Mallilinatham, Tamil Nadu	sugarcane
101	8861		4336	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Mallilinatham, Tamil Nadu	sugarcane
102	8862	6806	4344	*A*	*N. intermedia* (yellow)	*N. intermedia* (NiB)	East Asia	Ban Phru Mao, Thailand	sawdust compost, unburnt
103	8863		4358	*a*	*N. crassa*	*N. crassa* (NcC)	India	Mallilinatham, Tamil Nadu	sugarcane
104	8864		4359	*a*	*N. crassa*	*N. crassa* (NcC)	India	Madaurai, Tamil Nadu	sugarcane
105	8865		4360	*A*	*N. crassa*	*N. crassa* (NcC)	India	Madaurai, Tamil Nadu	sugarcane
106	8866		4361	*a*	*N. crassa*	*N. crassa* (NcC)	India	Rameshwaram, Tamil Nadu	sugarcane
107	8867		4362	*A*	*N. crassa*	*N. crassa* (NcC)	India	Rameshwaram, Tamil Nadu	sugarcane
108	8868		4363	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
109	8869		4364	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Madaurai, Tamil Nadu	sugarcane
110	8870	7833	4448	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
111	8871	7834	4449	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
112	8872		4453	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
113	8873		4454	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
114	8874		4464	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
115	8875		4480	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
116	8876		4481	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
117	8877		4490	*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
118	8878		4491	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
119	8879		4500	*a*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Franklin, Louisiana	sugarcane
120	8880	7847	4576	*A*	putative hybrid	PS2	Africa	Nosy Be, Madagascar	dry leaf material, unburnt
121	8881	7848	4578	*a*	putative hybrid	PS2	Africa	Nosy Be, Madagascar	dry leaf material, unburnt
122	8882	1543[6]	8045	*A*	*N. intermedia*	*N. intermedia* (NiA)	Carib. Basin	Puerto Cortes, Honduras	maize cob, burnt
123	8883	2009[6]	8064	*A*	*N. sitophila*	*N. sitophila*	Africa	Nigeria	soil isolation, unburnt

APPENDIX. Continued.

Strain numbers[1]				Mating type	Original species[2]	Phylogenetic species[3]	Geographic location		Substrate[5]
D	FGSC	Old FGSC	Other				Region[4]	Collection site	
124	8884	1843[6]	8085	*A*	*N. sitophila*	*N. sitophila*	Carib. Basin	Virginia	media contaminant
125	8885		M10	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
126	8886		M101	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
127	8887		M105	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
128	8888		M110	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
129	8889		M14	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
130	8890		M17	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
131	8891		M19	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
132	8892		M2	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
133	8893		M64	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
134	8894		M65	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
135	8895		M71	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
136	8896		M77	*a*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
137	8897		M86	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
138	8898		M92	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
139	8899		M95	*A*	*N. intermedia*	*N. intermedia* (NiA)	India	Maddur, Karnataka	sugarcane
140	8900	430[6]		*A*	*N. crassa*	*N. crassa* (NcA)	Africa	Adiopodoume, Ivory Coast	unknown
141	8901	434		*A*	*N. intermedia*	*N. intermedia* (NiA)	Africa	Monrovia, Liberia	unknown
142	8902	435[6]		*A*	*N. intermedia*	*N. intermedia* (NiA)	East Asia	Levuka, Fiji	unknown
143	8903	987[6]		*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Marrero, Louisiana	unknown
144	8904	1131[6]		*A*	*N. crassa*	*N. crassa* (NcA)	Carib. Basin	Cristobal, Panama	arthropod (mite)
145	8905	1270[6]		*A*	*N. tetrasperma*	*N. tetrasperma*	—	unknown	unknown
146	8906		W766	*a*	*N. discreta*	*N. discreta*	Carib. Basin	Belen, New Mexico	cottonwood tree
147	8907		W768	*A*	*N. sitophila*	*N. sitophila*	Carib. Basin	Belen, New Mexico	cottonwood tree

[1] Cross reference of strain numbers from different collections. Consecutive D numbers were assigned as convenient labels for the comparison of phylogenetic species recognition (PSR) and biological species recognition (BSR) (see Dettman et al. 2003). The entire collection was deposited in the Fungal Genetics Stock Center (FGSC). Some of the progenitors of these strains, prior to single conidium subculturing, were previously deposited in FGSC (Old FGSC). Other numbers (Other) refer to progenitors in the Perkins collection (now curated by FGSC), except M or W numbers, which refer to the personal Jacobson collection.

[2] Species designation originally assigned by traditional BSR (Perkins et al. 1976; Perkins and Turner 1988; Turner et al. 2001). Putative hybrid, strain originally designated a possible hybrid between *N. crassa* and *N. intermedia*. A question mark indicates species identification was regarded as questionable. The progenitors of strains D20, D22, and D38 were listed as *N. crassa*, and D72 as *N. intermedia*, in the Perkins database. Based on preliminary crosses with testers strains, these identifications were proven incorrect and classification was adjusted accordingly.

[3] PS1–3, Phylogenetic Species 1–3. Intraspecific subgroups in parentheses.

[4] Carib. Basin, Caribbean Basin; includes the coastal areas along the Gulf of Mexico and Caribbean Sea and the islands within. East Asia includes east of India and the Pacific islands.

[5] Vegetation was burnt, unless otherwise noted.

[6] These strains were used in previous phylogenetic analyses, including Natvig et al. 1987, Taylor and Natvig 1989, Randall and Metzenberg 1995, Skupski et al. 1997, Pöggeler 1999, and Dettman et al. 2001.