Species-Specific and Mating Type-Specific DNA Regions Adjacent to Mating Type Idiomorphs in the Genus Neurospora

Thomas A. Randall[1] and Robert L. Metzenberg

Department of Biomolecular Chemistry, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received December 19, 1994
Accepted for publication May 18, 1995

ABSTRACT

Mating type idiomorphs control mating and subsequent sexual development in *Neurospora crassa* and were previously shown to be well conserved in other Neurospora species. The centromere-proximal flanks of the *A* and *a* idiomorphs, but not the distal flanks from representative heterothallic, pseudohomothallic, and homothallic Neurospora species contain apparent species-specific and/or mating type-specific sequences adjacent to the well-conserved idiomorphs. The variable flank is bordered by regions that are highly homologous in all species. The sequence of ~1 kb immediately flanking the conserved idiomorphs of each species was determined. Sequence identity between species ranged from 20% (essentially unrelated) to >90%. By contrast, the *mt-A1* gene shows 88–98% identity. Sequence and hybridization data also show that the centromere-proximal flanks are very different between the two mating types for *N. intermedia*, *N. discreta*, and *N. tetrasperma*, but not for *N. sitophila* and *N. crassa*. The data suggest a close evolutionary relationship between several of the species; this is supported by phylogenetic analysis of their respective *mt-A1* genes. The origin of the variable regions adjacent to the evolutionarily conserved mating type idiomorphs is unknown.

WORK on the mating type system in the heterothallic fungus *Neurospora crassa* has advanced our understanding of sexual development in related filamentous ascomycetous fungi (GLASS and KULDAU 1992). *N. crassa* exists in nature exclusively in a haploid state as one of two stable mating types, *A* or *a*. A mating between the two types results in linear asci containing eight haploid, homokaryotic ascospores. Mating and sexual development in *N. crassa* are regulated by alternative *A* or *a* mating type idiomorphs located at allelic positions on linkage group I (GLASS *et al.* 1990a; STABEN and YANOFSKY 1990). These regions are termed idiomorphs because they have no significant sequence similarity yet are flanked on both sides by regions of DNA that are nearly identical at the sequence level. Sequence analysis has shown that the *A* idiomorph extends >5.3 kb and contains at least three open reading frames (GLASS *et al.* 1990a; GLASS and LEE 1992). The *mt-A1* ORF is the major regulator of mating in *N. crassa A* strains. Frameshift mutations within this ORF result in complete sterility (GRIFFITHS 1982; GLASS *et al.* 1990a). The *a* idiomorph extends >3.3 kb and contains one obvious ORF (*mt-a1*), which is the major regulator of mating in *N. crassa a* strains (STABEN and YANOFSKY 1990). Mating type idiomorphs with homology to *mt-A1* have been isolated from the heterothallic ascomycetes *Podospora anserina* (DEBUCHY and COPPIN 1992) and *Cochliobolus heterostrophus* (TURGEON *et al.* 1993). The subsequent characterization of both mating type idiomorphs from these fungi suggests that a system of idiomorphs like that seen in *N. crassa* is quite common in filamentous ascomycetes.

The genus Neurospora contains species exhibiting all three of the basic mating strategies employed by filamentous fungi: heterothallism, homothallism, and pseudohomothallism (secondary homothallism). Identification of heterothallic species in the genus Neurospora is based on both classical mycological criteria and on ability to mate productively with known tester strains (PERKINS *et al.* 1976). Sexual isolation in this genus is not absolute, as some interspecies matings will, with a very low frequency, result in production of viable ascospores. There are four known heterothallic Neurospora species, *N. crassa*, *N. sitophila*, *N. intermedia*, and *N. discreta*. The *A* and *a* mating types of each of these species contain sequences closely related to the mating type idiomorphs of *N. crassa*, as judged by DNA hybridization analysis (GLASS *et al.* 1990b). The mating type loci of *N. sitophila* and *N. intermedia* are known to be linked to genes whose functional counterparts in *N. crassa* are also on the mating type chromosome (linkage group I) (PERKINS *et al.* 1976; PERKINS 1977). No information is available for linkage relations in *N. discreta* (PERKINS and RAJU 1986).

N. tetrasperma is the only known pseudohomothallic species in the genus. *N. tetrasperma* is typically found as an *A*/*a* heterokaryon in which the eight progeny nuclei of the first postmeiotic division are packaged into only four spores, each containing both *A* and *a* nuclei. Upon

Corresponding author: Robert L. Metzenberg, Department of Biomolecular Chemistry, University of Wisconsin, Madison, WI 53706. E-mail: bobmetz@macc.wisc.edu

[1] *Present address:* Department of Genetics, Box 357360, University of Washington, Seattle, WA 98195.

germination, these spores regenerate a self-fertile mycelium. However, homokaryotic self-sterile *A* and *a* cultures can be resolved from self-fertile isolates (RAJU 1992). The mating type idiomorphs in *N. tetrasperma* are linked to genes that are analogs of those in *N. crassa*, and the idiomorphs are fully functional in *N. crassa* (METZENBERG and AHLGREN 1973; JACOBSON 1992).

The true homothallic species are numerous, but they can be subdivided into two groups. Members of the first group, consisting of *N. terricola* and *N. pannonica*, have sequences related to both *A* and *a* idiomorphs (GLASS *et al.* 1990b; T. A. RANDALL and R. L. METZENBERG, unpublished results). In the other group of homothallic species (including *N. africana*) only the *A* idiomorph has been detected; it is not known whether a distinct idiomorph that is functionally analogous to the *a* idiomorph but unrelated in sequence is also present or if the sexual cycle in these homothallic species is controlled by a single idiomorph. In all of the homothallic species, meiosis produces a linear eight spored ascus in which all progeny are self-fertile (RAJU 1977).

Preliminary studies of the centromere-proximal region flanking the two mating type idiomorphs in *N. crassa* showed a high degree of sequence similarity (>98%) (GLASS *et al.* 1990a; STABEN and YANOFSKY 1990). More extensive sequence analysis has shown that both flanks, and particularly the centromere-proximal flank, have minor differences that result in RFLPs (T. A. RANDALL and R. L. METZENBERG, unpublished results). In addition, the centromere-proximal flanking region of each mating type (within 4 kb of the idiomorphs) gives rise to a pair of RNA transcripts which show mating type-specific size differences. The transcript closest to the idiomorphs is 1.0 kb in the *A* mating type (and 1.4 kb in the *a* mating type) whereas the transcript adjacent to it is 1.5 kb in the *A* mating type (1.4 kb in the *a* mating type) (T. A. RANDALL and R. L. METZENBERG, unpublished results). This suggests that functionally significant differences between the mating types may occur outside the idiomorphs. The purpose of the present study is to examine in more detail the centromere-proximal flank within the known Neurospora species. The extensive differences revealed suggest the presence within this genus of a variable region immediately adjacent to the well conserved idiomorphs. This variable region is, in large part, species-specific and is, in several species, mating type-specific as well.

MATERIALS AND METHODS

Strains and basic molecular techniques: Strains used in this study are listed in Table 1. DNA was isolated essentially as described by STEVENS and METZENBERG (1982). Basic molecular techniques were performed as described in MANIATIS *et al.* 1982. DNA blots were made using Zetabind filters; all blots were probed with DNA fragments labeled with α-^{32}P-dCTP using the Decaprime labeling kit (Ambion Inc., Austin, TX) and washed under high stringency conditions (0.1× SSPE, 0.5% SDS at 65°C for ≥1 hr).

Cloning of *mt-A1* genes and adjacent variable regions from Neurospora species: Genomic clones were isolated from size-selected minilibraries of DNA digested with a single restriction enzyme and cloned into an appropriate pGEM vector (Promega, Madison WI). Approximately 1000–2000 *Escherichia coli* colonies were screened for each species with a probe from the *mt-A1* gene of *N. crassa* (*Kpn*I-*Kpn*I; nt 4025–4728 according to GLASS *et al.* 1990a). The *A* strains from which clones were isolated (Table 1) include the following: *N. crassa* OR23-IV *A* (L42302); *N. intermedia* Java T522-1 *A* (L42308); *N. tetrasperma* 85 *A* (L42310); *N. discreta* Kirbyville *A* (L42307); *N. sitophila* P8085 *A* (L42309); *N. sitophila* Arlington *A* and *N. africana* N200 (L42301). The *a* strains from which PCR products were amplified include the following: *N. crassa* ORS *a* (L42303); *N. intermedia* Java T522-8 *a* (L42305); *N. tetrasperma* 85 *a* (L42306); *N. discreta* Kirbyville *a* (L42304); and *N. sitophila* P8086 *a* (L42321).

Isolation of PCR products: PCR products were amplified from total genomic DNA under the following regime: denaturation at 94°C, 5 min; annealing at 55°C, 2 min; extension at 72°C, 3 min; followed by 29 cycles of 94°C, 1 min; 55°C, 2 min; 72°C, 3 min. PCR products were isolated from 1% agarose gels using the Qiaex system (Qiagen Inc., Chatsworth, CA), dissolved in TE, pH 8.0 and used for sequencing without further purification.

DNA sequencing and strategy: DNA sequencing from plasmid templates was done using the dideoxy procedure and Sequenase version 2.0 from USB, Cleveland, OH. Sequencing of PCR products was with the fmol DNA Sequencing system (Promega, Madison, WI). Primers used in DNA sequencing and/or PCR were synthesized by either the University of Wisconsin Biotechnology Center or the Marshall University DNA Core Facility, Huntington, WV (Table 2). The *mt-A1* genes from all species were sequenced using primers 1778, 2043, 1874, 1875, and 3194 (Table 2). The primers used to sequence the variable regions are listed in the linear order used, in Table 2, beginning from the 3′ end of the *mt-A1* ORF in each species and extending toward their respective variable regions. All DNA sequence was obtained from the clones and PCR fragments described in Figure 3 and below. Subclones from pntAF [pTF2 (0.8-kb *Bam*HI-*Hin*dIII) and pTF3 (1.5-kb *Bam*HI-*Hin*dIII)], pniF2 [pIF2 (0.7-kb *Nsi*I-*Eco*RI) and pIF3 (1.3-kb *Eco*RI-*Sac*I)] and pndCla2 [pndNX (0.9-kb *Nsi*I-*Xho*I) and pndH3 (1.0-kb *Hin*dIII)] were sequenced using forward and reverse primers. All DNA sequence comparison was performed using LASERGENE software from DNASTAR Inc., Madison, WI. Pairwise comparisons between individual sequences were made using the Martinez/Needleman-Wunsch algorithm with default settings as follows: minimum match = 9; gap penalty = 1.1; and gap length penalty = 0.33. With this algorithm, the percent DNA sequence similarity is defined as number of matching residues/mismatches + matching residues + gaps × 100. We will use the term sequence similarity when we are discussing DNA sequence comparisons; the term homology will be used only when results from Southern hybridization analysis are being discussed (because an absolute value of similarity cannot be measured in those analyses). Variable regions with DNA sequence similarity in the range of 20–25% will be referred to as unrelated [because we consider such sequences to have no more obvious relation to each other than to randomly generated sequences of identical base-pair composition (data not shown)]. The percent similarities listed throughout have all been rounded off to the nearest integer. All DNA sequences reported here are being deposited in Genbank.

RESULTS

DNA hybridization analysis of the heterothallic and pseudohomothallic Neurospora species: In this study,

TABLE 1

Strains used in this study

Species	Strain	Mating type	Source
N. crassa	OR23-IV	*A*	FGSC 2489
N. crassa	Mauriceville-1c	*A*	FGSC 2225
N. crassa	ORS	*a*	FGSC 2490
N. discreta	Kirbyville-6	*A*	FGSC 3228
N. discreta	Kirbyville-1	*a*	FGSC 3229
N. discreta	Sogeri Road-1	*A*	FGSC 6786
N. discreta	Wau-6	*a*	FGSC 6784
N. intermedia	Java T522-1	*A*	BCT, SG# 8139
N. intermedia	Java T522-8	*a*	BCT, SG# 8140
N. intermedia	Rouna P758	*A*	BCT (RLM 57-13)
N. intermedia	Rouna P757	*a*	BCT (RLM 55-29)
N. intermedia	Taipei-1 P13	*A*	FGSC 1766
N. intermedia	Taipei P17	*a*	FGSC 1767
N. sitophila	Arlington	*A*	FGSC 417
N. sitophila	Europe-2	*a*	FGSC 1779
N. sitophila	P8085	*A*	FGSC 2216
N. sitophila	P8086	*a*	FGSC 2217
N. tetrasperma	85	*A*	FGSC 1270
N. tetrasperma	85	*a*	FGSC 1271
N. tetrasperma	Gianjor-1	*A*	FGSC 1794[a]
N. tetrasperma	Gianjor-1	*a*	FGSC 1794[a]
N. tetrasperma	Groveland-1	*A*	FGSC 1942[a]
N. tetrasperma	Groveland-1	*a*	FGSC 1942[a]
N. tetrasperma	Liberia	*A*	FGSC 965[a]
N. tetrasperma	Liberia	*a*	FGSC 965[a]
N. tetrasperma	Homestead-1f	*A*	FGSC 1943[a]
N. tetrasperma	Homestead-1f	*a*	FGSC 1943[a]
N. tetrasperma	Labelle-1	*A*	FGSC 1941[a]
N. tetrasperma	Welsh-1 bv-1	*A*	FGSC 2503
N. tetrasperma	Welsh-1 bv-1	*a*	FGSC 2504
N. tetrasperma	Perkins IV-2	*A*	FGSC 2505
N. tetrasperma	Perkins IV-2	*a*	FGSC 2506
N. tetrasperma	Lihue	*A*	FGSC 2509
N. tetrasperma	Hanalei-1 bv-1	*A*	FGSC 2510
N. tetrasperma	Hanalei-1 bv-3	*a*	FGSC 2511
N. tetrasperma	Raleigh	*A*	FGSC 3998
N. tetrasperma	Raleigh	*a*	FGSC 4245
N. tetrasperma	Empire-1 V-2	*A*	FGSC 2501
N. tetrasperma	Empire-1 V-2	*a*	FGSC 2502
N. tetrasperma	Ahipara N.Z. P4371	*A*	DJJ
N. tetrasperma	Ahipara N.Z. P4372	*a*	DJJ
N. tetrasperma	Ahipara N.Z. P4383	*A*	DJJ
N. tetrasperma	Ahipara N.Z. P4384	*a*	DJJ
N. tetrasperma	Franklin LA W1	*A*	DJJ
N. tetrasperma	Franklin LA W2	*a*	DJJ
N. tetrasperma	Franklin LA W23	*A*	DJJ
N. tetrasperma	Franklin LA W24	*a*	DJJ
N. tetrasperma	Franklin LA W27	*A*	DJJ
N. tetrasperma	Franklin LA W28	*a*	DJJ
N. africana	N200		FGSC 1740
N. lineolata	A-236		FGSC 1910
N. dodgei	PR-300		FGSC 1692
N. galapagosensis	G-349		FGSC 1739

BCT, BARBARA C. TURNER; SG, TURNER's silica gel number; DJJ, DAVID J. JACOBSON; RLM, ROBERT L. METZENBERG stock number, where TURNER's silica gel number is not known.

[a] Single mating type isolates were obtained from the indicated FGSC strains.

TABLE 2

Primers

Primer	Sequence
A. Primers and their sequence	
Forward	CGCCAGGGTTTTCCCAGTCACGAC
Reverse	TCACACAGGAAACAGCTATGAC
1778	TCCAGCTGCACCCAAACTTCCCACC
2043	GTTCGCCGAATCCCCGC
1874	CTCGAGGTCGTGAGTGC
1875	TGTATTCGTCAATCCGG
3194	ATGGATCCTCATCTTCCACTAACCC
a7	TAACTCGGTAACCGT
a8	GCGAGAGTAGAATGG
a10	TGCGTATCTGGCAGAAT
a16	CAGAACTATATTCCC
a17	AGGTACGGTGACAAC
a18	CCTAATTGCGGAGT
a20	ACCGAGGTAAATGAG
a32	CAGTTTGAGACCTTGTTGG
a33	CAGGTCTGGACCACG
a36	GGATTTAACGGAGGAGC
a37	CGGGGTGGAATGGGC
a38	ACAGGGTTGCTCTGC
a39	TGCGGGAGCTGTGGC
a44	GCAGTCAAGTCCCG
a45	CGCTGACGCCAACGC
a51	CACTAAGCAAACCC
a53	GACTCGCGAGTCGC
a58	CAGTAGAAGCTGTGCC
a59	GGACAGCTGCCTCCC
a61	CAGTACACCTTTGCG
a62	TCTACGAGTCCGGG
a63	TTACTAAATCTAGGC
a64	CATTGGTTGATCGG

Strain	Primers
B. Linear order[a] of primers for each strain	
N. tetrasperma A	a32, a36, a37, a45, a58, a59
N. intermedia A	a32, a36, a37, a45, a58
N. discreta A	a32, a36, a37, a59, a64
N. sitophila A	a32, a36, a37, a38, a62, a7, a8
N. africana	a32, a36, a37, a61
N. tetrasperma a	a33, a20, a38, a39, a44, a51
N. intermedia a	a33, a20, a38, a39, a44, a51
N. discreta a	a33, a20, a39, a53, a60
N. sitophila a	a33, a20, a38, a39, a62, a7, a8

[a] Linear order means that in the case of *N. tetrasperma A*, for instance, the sequence obtained with primer a36 overlaps with that of a32, etc. Primer a32 or a33 is within the 3′ end of the *A* or *a* idiomorph, respectively, of each strain and the sequence of variable regions analyzed in the present study can be obtained by linking the sequence obtained from one primer to the previous primer with the subclones described below.

we considered three regions of interest: (1) variable regions, which are immediately centromere-proximal to the idiomorphs and common regions, and are completely or almost completely dissimilar between species; (2) idiomorphs, which are highly conserved between species but are completely dissimilar between the two mating types within the species (excluding *N. africana* and its congeners; see Introduction); (3) a "mating type-common region" of 57–59 bp, which is highly similar between all species (including the self-fertile *N. africana*) and between both mating types within each species in which there are two mating types; it separates an idiomorph from its nearby variable region. In addition to the 57–59-bp region that is common to both mating types, there is an additional region of ~140 bp that is common to all *a* mating type strains.

Initial molecular characterization of mating types in the Neurospora spp. had been limited to the identification of the presence or absence of idiomorphs by DNA hybridization (GLASS *et al.* 1990b). This has now been extended to the flanking regions. Southern blots containing DNA from all species were hybridized with probes 1–8 from *N. crassa A*. A schematic summary of the results obtained with these eight *N. crassa A* probes is given in Figure 1A. A probe from the centromere-distal side of the *N. crassa A* idiomorph (probe 1 of Figure 1A) showed homology to all species, including *N. africana* (data not shown) (GLASS and SMITH 1994). By contrast, a probe from the centromere-proximal idiomorph flank of the *A* idiomorph of *N. crassa* (probe 4 of Figure 1A) showed both species-specific and mating type-specific hybridization. A representative result is shown in Figure 1B. Results using selected probes from the variable regions of other species and the locations of these regions are outlined here for their value as examples, and are discussed more fully in the APPENDIX.

We wished to know whether the mating type-specific differences of these variable regions could be seen in other independently collected isolates of these species. We therefore chose a set of *N. intermedia* and *N. tetrasperma* isolates of geographically diverse origins and prepared separate Southern blots containing genomic digests from six strains of *N. intermedia* (three of each mating type) and from four pairs of the resolved (homokaryotic) mating types of *N. tetrasperma*. The *N. intermedia* blot was probed sequentially with DNA prepared from the following fragments of the variable region of *N. intermedia a*: the 1.5-kb *Eco*RI-*Sac*I fragment (insert from pIF3; box B in Figures 1A and 3A) of the variable region of *N. intermedia A* (Figure 1C) and with a 550-bp *Spe*I-*Spe*I fragment (box C of Figures 1A and 3B; data not shown). With each probe, the presence or absence of hybridization was mating type-specific, *i.e.*, DNA homologous to the appropriate cloned region of the variable region of the *A* or *a* mating types under consideration was present in the appropriate mating type of all isolates examined and was absent from the other mating type. Similar results were observed with the *N. tetrasperma* and *N. discreta* strains (data not shown).

In Figure 1A, each of the probes used is represented as a box with a different fill-pattern. Because *N. sitophila A* and *a* have homology to all *N. crassa A* probes, we tentatively assume that these flanks are in the same

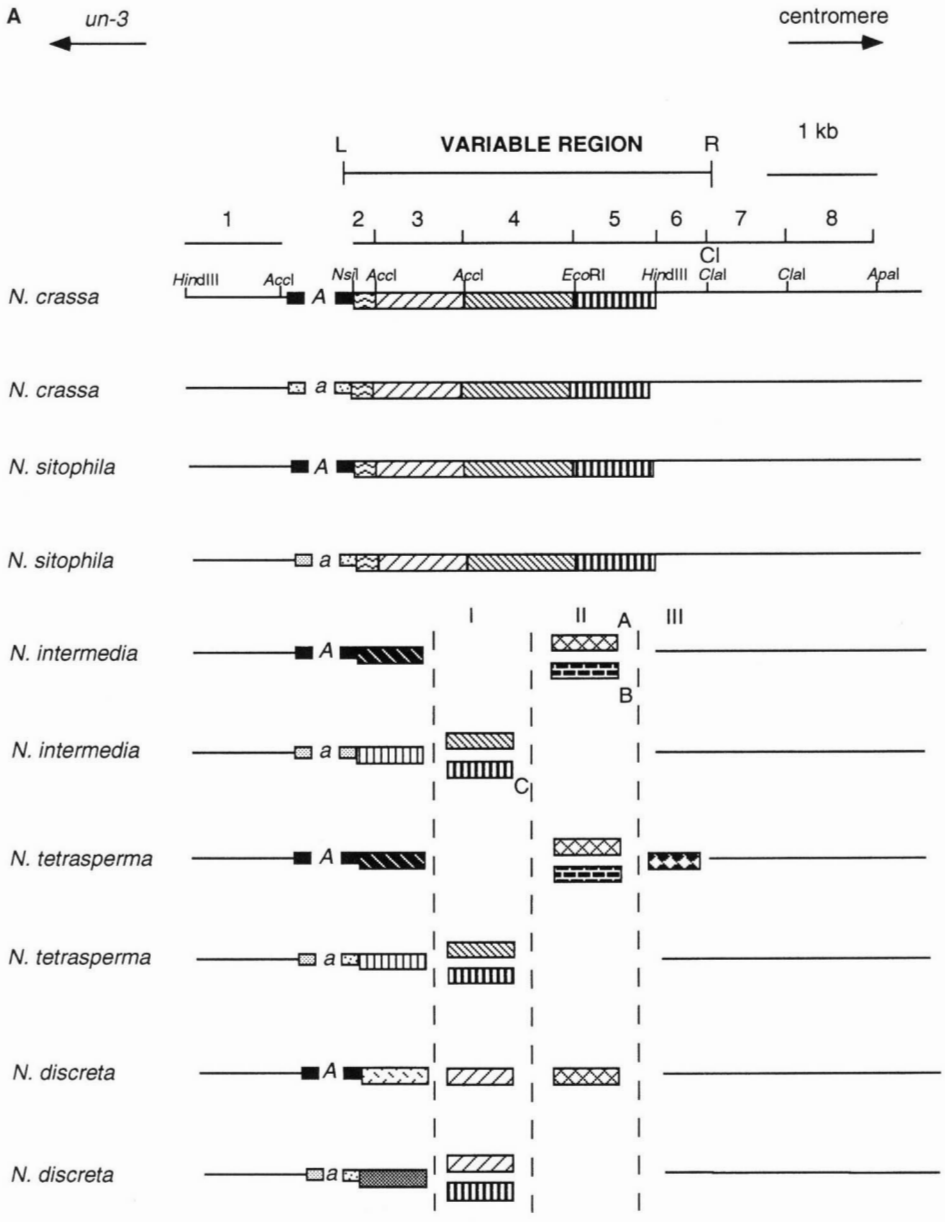

FIGURE 1.

physical order as those of *N. crassa*. The floating boxes in column I indicate hybridization to the corresponding probes from the *N. crassa A* strain, but that no assumption about their order is warranted. In many cases, the probe from *N. crassa* showed no hybridization, indicating that these species completely lack a corresponding homologous region of DNA in their genomes. The absence of a particular box in column I indicates no hybridization to that strain. Where a fragment of DNA from the flank of one of the other species was used as a probe (for example, boxes labeled A and B), a new characteristic fill-pattern is used to indicate the species to which that probe hybridizes; cross-hybridization to other species with these non-*N. crassa* probes is also indicated with floating boxes (column II). Column III indicates a DNA fragment from *N. tetrasperma A* that does not hybridize to DNA of any of the other species.

We conclude that these results are representative for each particular species because all strains tested gave identical results (Table 1); Figure 1A is a summary based on all these strains. We will use the word "variable region" to refer to this region of DNA bordered by a mating type common region (see below) adjacent to a conserved idiomorph on one side and bordered on the other side by conserved DNA sequences present in all heterothallic and pseudohomothallic species examined (thin black lines in Figure 1A; data from Southern analysis, not shown). By our definition, a variable region for each species can contain DNA that may be present in some other species, but not all; present in one mating type of several species but not the other; or unique to one mating type of a particular species. This is in contrast to idiomorphs, which differ radically between mating types of the same species but are well conserved

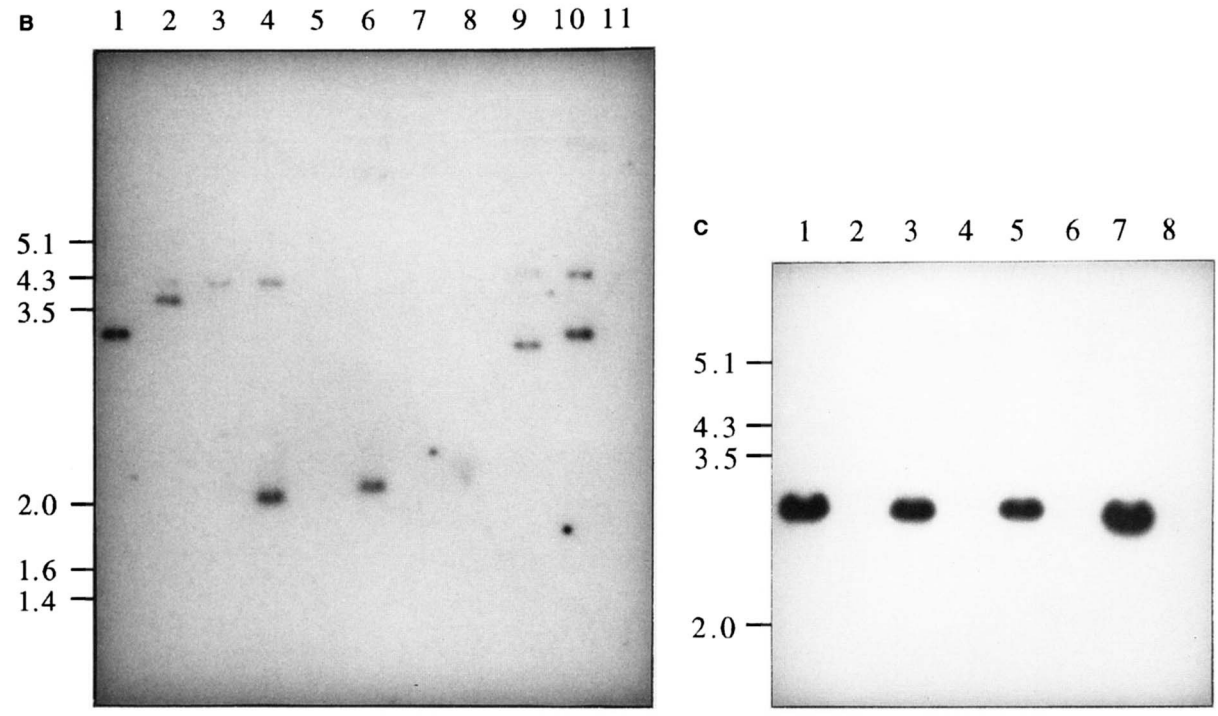

FIGURE 1.—*Continued*—Hybridization differences in the regions centromere-proximal to the mating type idiomorphs: a summary of the variable regions of all heterothallic Neurospora species. (A) Probes 1–8, indicated above the *N. crassa A* line drawing were used to probe blots like that shown in B. The narrow boxed regions [solid black fill-pattern (*A*) and dotted (*a*)] indicate the idiomorphs: these are not to scale. Thin continuous lines indicate that hybridization occurred with the corresponding *N. crassa A* DNA probe (at high stringency). Probes 2–5 are represented by distinct fill-patterns; hybridization to these *N. crassa A* probes is indicated by the presence of the appropriate fill pattern attached to the idiomorph or by a floating box in column I. Those boxes that are attached to either idiomorph in the line drawings representing *N. intermedia*, *N. tetrasperma*, and *N. discreta* indicate regions that have been sequenced (see below) and therefore have been shown to be physically adjacent to the idiomorphs. Those boxes that appear in A as floating boxes (unattached to an idiomorph) are represented as such because we are making no presumption as to the relative position of the various hybridizing regions. However, we note that they are within the cloned regions or PCR fragments discussed below, and, therefore, are physically linked to the idiomorphs. Fragments of DNA from other species that were used as DNA probes were given different fill-patterns and hybridization in any particular species is shown in column II (see Figure 3 for origin of each probe). Column III is a region that appears to be unique to *N. tetrasperma A*, because no hybridization to probe 6 was observed. The Cl above the *Cla*I site indicates the point beyond which hybridization is seen to all strains in this figure using probes from *N. crassa A*; this is therefore considered the approximate location of the centromere-proximal border of the variable regions. The L and R indicate the left and right ends of the variable region. (B) Southern blot of various Neurospora species. *Eco*RI digests of genomic DNAs were run on a 1% agarose gel and transferred to a Zetabind membrane. The blot was probed with fragment 4 of A. The sizes of markers are in kilobases. Lane 1, *N. crassa* 74-OR23-IV *A*; lane 2, *N. crassa* ORS *a*; lane 3, *N. tetrasperma* 85 *A*; lane 4, *N. tetrasperma* 85 *a*; lane 5, *N. intermedia* Java *A*; lane 6, *N. intermedia* Java *a*; lane 7, *N. discreta* Kirbyville *A*; lane 8, *N. discreta* Kirbyville *a*; lane 9, *N. sitophila* P8085 *A*; lane 10, *N. sitophila* P8086 *a*; lane 11, *N. africana*. The faint band of hybridization seen at 4.3 kb in lane 3 was found not to correspond to the cloned regions representing the flanks of this species (see below). A faint band of what is probably the same sequence is also seen in lanes 1, 2, 4, 9 and 10. This, and some even weaker bands that could be detected on the autoradiogram but not on photographs, are not considered further in this study. (C) Identification of the variable regions in mating type pairs of various *N. intermedia* and *N. tetrasperma* isolates. A 1% agarose gel of *Eco*RI digests of genomic DNAs were transferred to a Zetabind membrane. Lane 1, *N. intermedia* Rouna *A*; lane 2, *N. intermedia* Rouna *a*; lane 3, *N. intermedia* Java *A*; lane 4, *N. intermedia* Java *a*; lane 5, *N. intermedia* Taipei *A*; lane 6, *N. intermedia* Taipei *a*; lane 7, *N. tetrasperma* 85 *A*; lane 8, *N. tetrasperma* 85 *a*. The probe was an *A*-specific DNA fragment within the variable region of *N. intermedia A* (box labeled B in Figure 1A, described in the text). The size of markers is in kilobases.

within each mating type in all species in the genus. The left border of the variable regions will be defined as the first base pair beyond the mating type or *a* common regions (see below). The right border of the variable regions will be defined as the point at which the sequences return to a high degree of similarity; this corresponds approximately to the *Cla*I site labeled C in Figure 1A. *N. crassa* probes from beyond this point (*viz.*, the *Cla*I-*Cla*I fragment designated probe 7) hybridize at high stringency to both mating types of all heterothallic species and of *N. tetrasperma* (data not shown). A more detailed discussion of the right borders of the variable regions is given in the APPENDIX.

DNA hybridization analysis of true homothallic Neurospora species: A separate analysis of those homothallic species containing only the *A* idiomorph was performed. Using the eight DNA probes from a centromere-proximal flanking region of *N. crassa A*, we

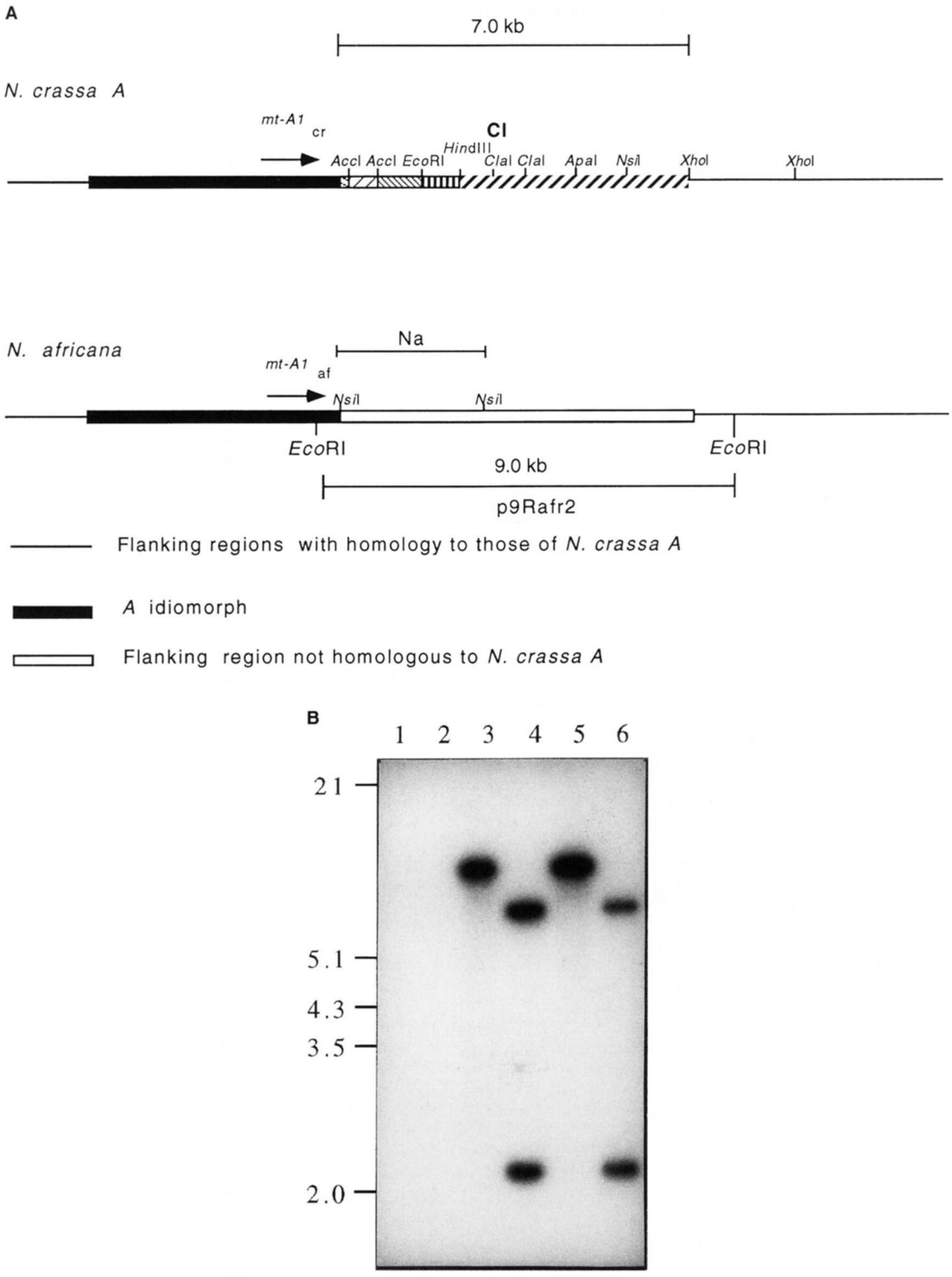

FIGURE 2.—Characterization of the variable region in homothallic species. (A) Comparison of *N. crassa A* with *N. africana*. Boxes with solid black fill indicate position of the *A* idiomorphs. Continuous thin lines indicate homology between *N. crassa* and *N. africana*. The region of the *N. crassa* chromosome with striped and herringbone fill-patterns indicates the extent of the *N. crassa* flank that showed no hybridization to genomic DNAs of the homothallic species at high stringency. The Cl above the *Cla*I site indicates the point beyond which hybridization is seen to all strains in Figure 1A using probes from *N. crassa A*; this is therefore considered the approximate location of the centromere-proximal border of the variable regions of the heterothallic and pseudo-homothallics strains. The 9-kb *Eco*RI fragment of *N. africana* that was cloned and analyzed is indicated. The line above the *N. africana* flank (Na) indicates the position of the DNA fragment used as a probe in B. (B) A 1% agarose gel of *Eco*RI digested genomic DNAs was transferred to a Zetabind membrane. The probe was the *Nsi*I-*Nsi*I fragment of *N. africana*. The sizes of markers are in kilobases. Lane 1, *N. crassa* 74-OR23-IV *A*; lane 2, *N. crassa* ORS *a*; lane 3, *N. africana*; lane 4, *N. dodgei*; lane 5, *N. galapagosensis*; lane 6, *N. lineolata*.

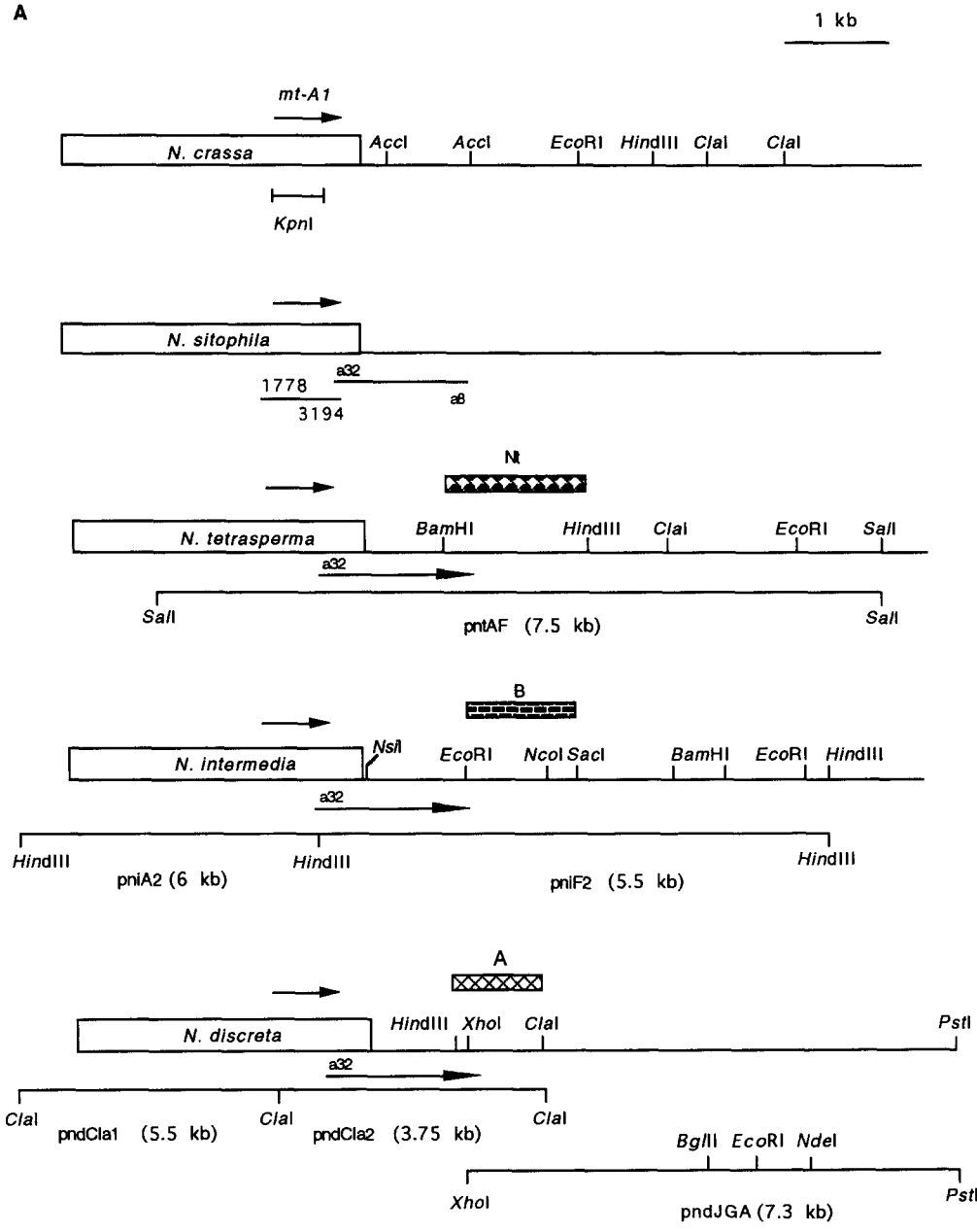

FIGURE 3.—Diagrams of plasmids and PCR fragments used for analysis. (A) Clones of variable regions flanking the *A* idiomorphs. A line drawing representing each variable region is shown attached to a box representing the *A* idiomorph. The filled boxes represent restriction fragments used as probes in Figures 1 and 7. The locations of the subclones isolated from minilibraries are listed below each line drawing. The positions of pntAF, pniA2, pniF2, pndCla1, and pndCla2 with respect to the *A* idiomorph were determined by sequencing of the borders of each insert with standard forward or reverse primers. The position of the *KpnI-KpnI* fragment used as a probe for the isolation of pntAF, pniA2, pniF2, pndCla1 and pndCla2 is shown below the *N. crassa* line drawing. The clone pndJGA was found by screening a minilibrary with the fragment designated A. Positions of the initial primer used in sequencing (a32) is indicated below each drawing and the arrow indicates the extent of sequence data obtained. The other primers used in sequencing each variable region are listed in Table 2A and are listed in the order used following primer a32 (Table 2B). The endpoints of all subclones were sequenced in order to determine the position of the subclones with respect to the *A* idiomorph. Subclones from pntAF (pTF2 and pTF3), pniF2 (pIF2 and pIF3), and pndCla2 were generated to complement the above strategy when appropriate see (MATERIALS AND METHODS). For *N. sitophila* the location of the PCR products used in sequencing is shown. The boxes labeled Nt, B, and A show the positions of DNA fragments used as probes in subsequent experiments. (B) PCR fragments of variable regions flanking the *a* idiomorphs. A line drawing of the *N. crassa a* flank is shown with the positions of the primers used to amplify corresponding regions of the other species (primers a8 and a10; a 2.1-kb fragment is amplified from *N. crassa a* with primer a10). The sizes of the PCR fragments amplified from each species and the primers used in their amplification are indicated. Symbols in filled boxes correspond to those in Figure 1A; the region of the *N. intermedia a* variable region labeled C was used as a DNA probe.

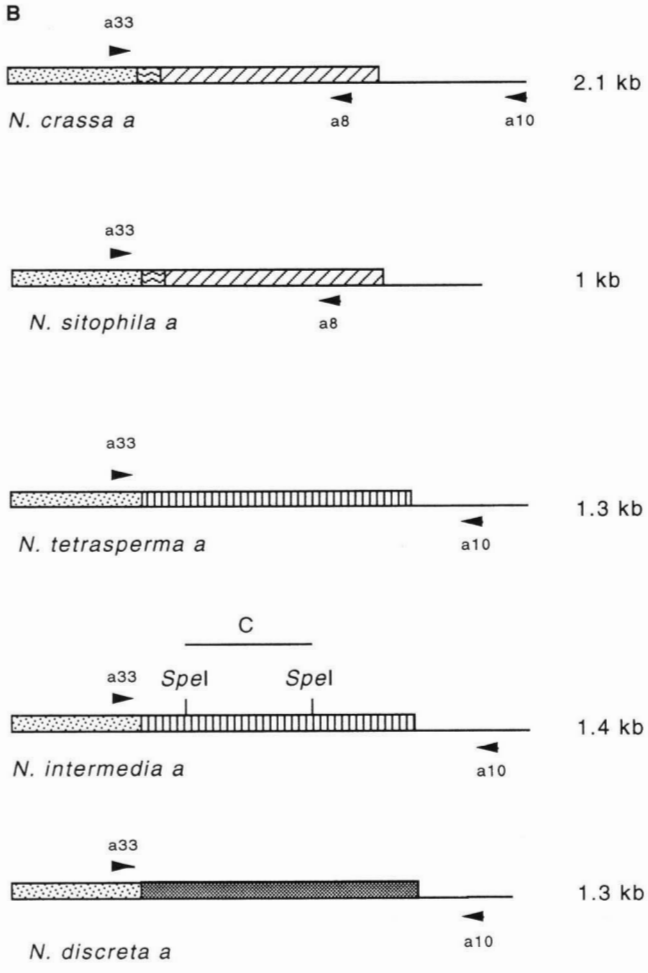

FIGURE 3.—*Continued*

found no hybridization to *N. africana*, *N. lineolata*, *N. dodgei*, or *N. galapagosensis* (see Figure 1A and summary in Figure 2A). A 9-kb *Eco*RI fragment containing this region of difference and the flanking homologous region was cloned from *N. africana*, our model homothallic species. Using a probe within this *N. africana* flank (Figure 2B, the *Nsi*I-*Nsi*I fragment of Figure 2A), we found that the corresponding DNA is present in all homothallic species tested but is not present in any other heterothallic or pseudohomothallic Neurospora species (see below). This indicated that the difference seen in this region is not explainable as a deletion in the homothallic species relative to *N. crassa* that has brought distant sequences closer to the mating type idiomorph, but reflects a region of DNA unique to these species. Thus, this region can be considered a homothallic-specific variable region. A comparison of Figures 2A and 1A shows that the variable region present in all these homothallic species extends an additional 3 kb relative to that of the heterothallic and pseudohomothallic species. *N. crassa A* DNA probes between the *Cla*I site (labeled C in both Figures 1A and 2A) and the *Xho*I site in *N. crassa A* (see Figure 2A) have homology to DNA of all of the heterothallic species and to *N. tetrasperma*, but lack homology with the homothallic species.

Sequence analysis of the variable regions of *A* strains: To determine more precisely the extent of divergence, we cloned the variable regions of the *A* mating type strain of each species (Figure 3A). Starting with a primer complementary to a region of suspected high sequence similarity in the 3′ ends of the *mt-A1* ORFs of all idiomorphs (primer a32, Figure 3A), we sequenced ~1300 bp from each species. A high level of similarity (>90%) was seen between all species from the 3′ end of the *mt-A1* gene to the border of the *A* idiomorph (point C_A, Figure 4A). This high level of similarity extended 59 bp beyond the idiomorph border (to point L_A; Figure 4A) and is designated the mating type common region; beyond the mating type common region, sequence similarity dropped from ~90–95% to 20–65% (Table 3).

The sequenced portion of the *N. crassa* and *N. sitophila* variable regions (designated the *crassa/sitophila* subgroup) have an overall level of sequence identity of 65%, but a closer examination shows that the differences are due mainly to a deletion in the variable region of *N. crassa* relative to that of *N. sitophila*. See the APPENDIX for a complete discussion of the sequence similarities between various variable regions and of the apparent insertions in the centromere-proximal flank of the *A* idiomorph of *N. crassa* shown in Figure 4A. The sequenced portions of the *N. intermedia* and *N. tetrasperma* variable regions (designated the *intermedia/tetrasperma* subgroup) also have a substantial level of sequence similarity (47%). There is no significant similarity between the variable regions of the members of the two subgroups. Between the variable regions of *N. africana*, *N. discreta*, and *N. intermedia* there is a region of ~400 bp that is 50–60% similar between all three species. Between various pairs of variable regions there are also smaller regions of 100–200 bp that have a significantly higher level of sequence similarity than do the variable regions taken as a whole (see APPENDIX). We will call these regions of significant similarity between any pair of variable regions "islands of homology".

Sequence analysis of the variable regions of *a* strains: Because the corresponding regions of difference between species in the *a* strains were somewhat smaller than those of the *A* strains, we used PCR and appropriate primers to amplify those DNAs (Figure 3B). Beginning at a position 68 bp inside the *a* idiomorph border we sequenced ~1000 bp of the variable region from each species. A high level of sequence similarity of the *a* idiomorph was observed between all species up to the *a* idiomorph border. Past the idiomorph border, all species maintained this level of similarity for approximately another 195 bp (designated the *a* common region, of which the first 57–59 bp is also seen centromere-proximal to the *A* idiomorph; see Figure 4B). The first base pair at which the flanks of the *A* and

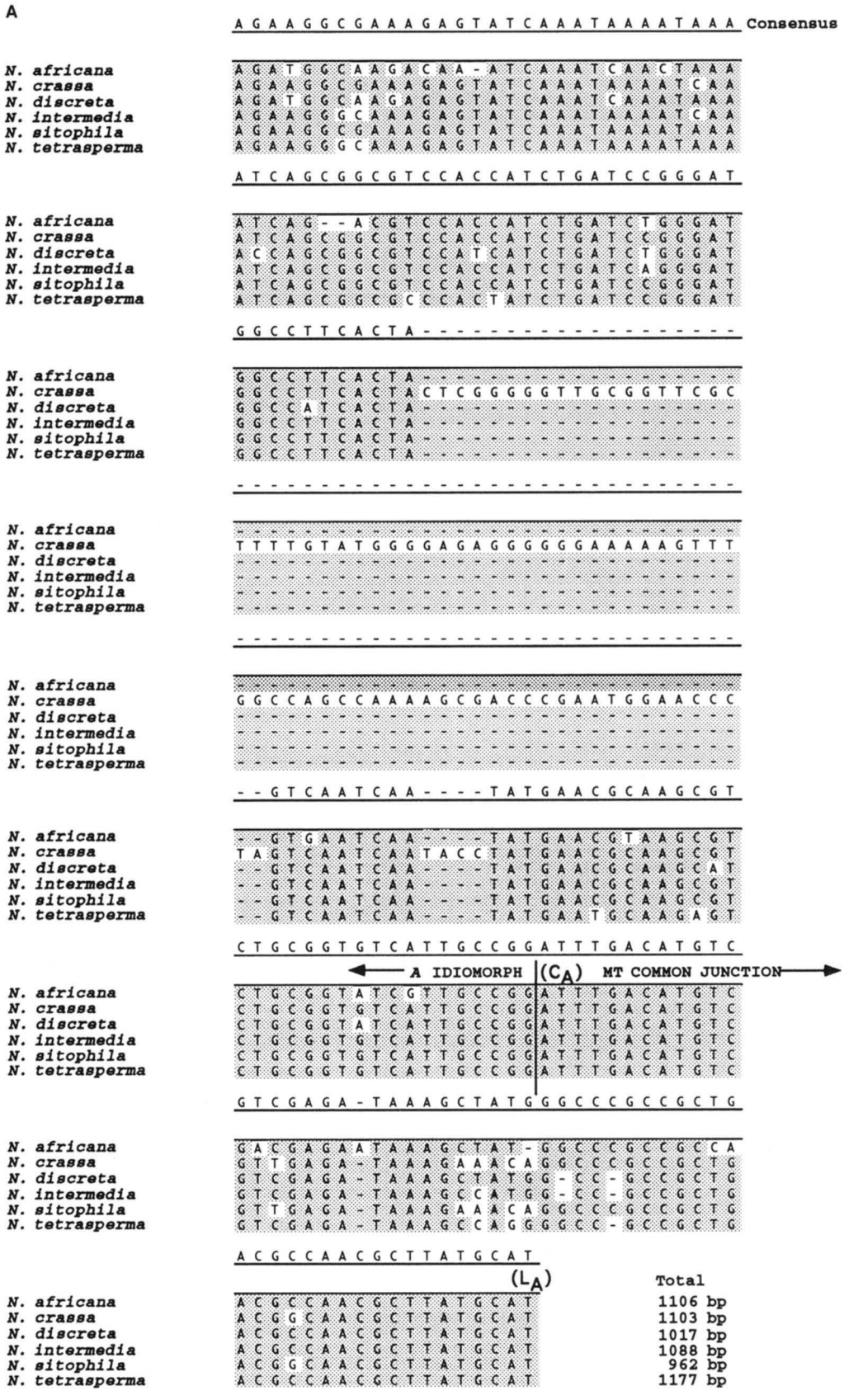
A
A G A A G G C G A A A G A G T A T C A A A T A A A A T A A A Consensus
N. africana
N. crassa
N. discreta
N. intermedia
N. sitophila
N. tetrasperma
A T C A G C G G C G T C C A C C A T C T G A T C C G G G A T
G G C C T T C A C T A
T T T T G T A T G G G G A G A G G G G G A A A A A G T T T
G G C C A G C C A A A A G C G A C C C G A A T G G A A C C C
- - G T C A A T C A A - - - - T A T G A A C G C A A G C G T
- - G T G A A T C A A - - - - T A T G A A C G T A A G C G T
T A G T C A A T C A A T A C C T A T G A A C G C A A G C G T
C T G C G G T G T C A T T G C C G G A T T T G A C A T G T C
A IDIOMORPH
(CA)
MT COMMON JUNCTION
G T C G A G A - T A A A G C T A T G G G C C C G C C G C T G
A C G C C A A C G C T T A T G C A T
(LA)
Total
1106 bp
1103 bp
1017 bp
1088 bp
962 bp
1177 bp

FIGURE 4.

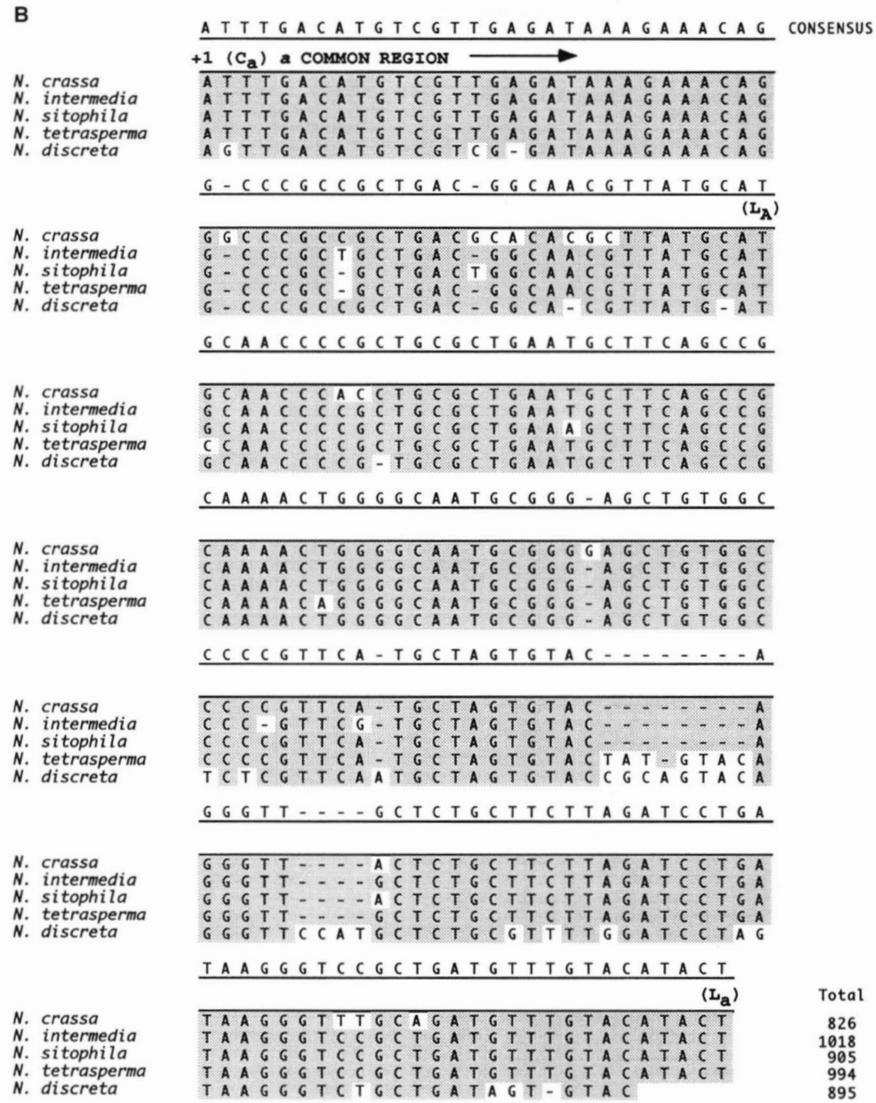

FIGURE 4.—*Continued*—Sequence comparison of the idiomorph and common region bordering the variable region. (A) Sequence from the region of sequence similarity within the *A* idiomorph toward its variable region. Dashes in the sequence indicate a gap relative to the *N. crassa* idiomorph sequence. Point C_A indicates the centromere-proximal border of the *A* idiomorph and point L_A indicates the left border of the *A* variable region; the region between C_A and L_A is designated the mating type common region (MT Common Junction in figure). The total amount of sequence obtained from each variable region (beginning at point L_A) is listed at the bottom of the figure. (B) Sequence comparison of the *a* idiomorph-variable region border. Only the sequence of the *a* common region is shown. Point C_a is the first nucleotide beyond the *a* idiomorph; point L_a is the centromere-proximal border of the *a* common region. Point L_A (the point at which the variable regions in the *A* mating types begin) is shown for comparison. The total amount of sequence obtained from each variable region (beginning at point L_a) is listed at the bottom of the figure.

a idiomorphs of *N. crassa* (and all other species; see below) resume sequence similarity was designated the first base pair of the *a* common region (C_a; identical to C_A). The 57–59 bp of the *a* common region, C_a to L_a, is identical to the region between C_A to L_A; this will be called the mating type common region.

Like the variable regions of the *A* strains, the variable regions of the *a* strains of *N. crassa* and *N. sitophila* have a high level of sequence similarity (from bp 1–900 of each variable region there is 90% similarity between the two species), as do the variable regions of *N. intermedia a* and *N. tetrasperma a* (71% similarity; Table 4). The sequenced regions of the variable regions of the *a* strains of the *crassa/sitophila* subgroup have only 25–34% similarity to those of the *intermedia/tetrasperma* subgroup; also, no islands of homology between the subgroups were observed. The variable region of *N. discreta a* appears unrelated to those of *N. crassa a* and *N. sitophila a*, but several islands of homology exist between this variable region and those of the *a* strains of the *intermedia/tetrasperma* subgroup. What would otherwise be the *a* common region of *N. discreta a* appears to have an insertion of 90 bp (at nucleotides 1–90) relative to the comparable regions of the other *a* strains, because

TABLE 3

Sequence comparison of variable regions of *A strains*

	Percent similarity[a]					
	1	2	3	4	5	6
N. africana (1)	—					
N. crassa (2)	25	—				
N. discreta (3)	43	24	—			
N. intermedia (4)	33	20	40	—		
N. sitophila (5)	22	65	25	20	—	
N. tetrasperma (6)	26	21	24	47	23	—

Calculations were made beginning from point L_A, the first base pair of each variable region.

[a] As defined in MATERIALS AND METHODS.

TABLE 4

Sequence comparison of variable regions of *a* strains

	Percent similarity				
	1	2	3	4	5
N. crassa (1)	—				
N. intermedia (2)	32	—			
N. sitophila (3)	90	34	—		
N. tetrasperma (4)	33	71	34	—	
N. discreta (5)	26	34	26	42	—

Calculations were made beginning from point L_a, the first base pair of each variable region.

bp 91–224 of this region of *N. discreta a* are identical to the region designated as the *a* common region for the other *a* mating types (Figure 4B). Within the regions sequenced in the variable regions of *N. discreta a*, *N. tetrasperma a* and *N. intermedia a*, there appear to be no islands of homology as large as that observed with the corresponding variable regions in the corresponding *A* strains.

Sequence comparison between the variable regions of the *A* and *a* strains of the same species: Previous sequence analysis of *N. crassa* clones defined the borders between the *A* and *a* idiomorphs and their flanking sequences as the point at which a high level of sequence similarity returned (GLASS et al. 1990a; STABEN and YANOFSKY 1990) (point C_A of Figure 4A and C_a of Figure 4B). Further sequencing of this region in *N. crassa A* and *a* showed that this high level of sequence similarity extends for ≥3 kb toward the centromere (overall >90%) (T. A. RANDALL and R. L. METZENBERG, unpublished results) and most likely includes the entire remainder of the chromosome. The available sequences of the variable regions of the *A* and *a* strains of the heterothallic species and of *N. tetrasperma* were compared; the results are summarized in Figure 5. Both *N. crassa* and *N. sitophila* show a high degree of sequence similarity between the variable regions of the two mating types (96% for *N. sitophila* and 66% for *N. crassa*). The 200-bp deletion (discussed above) in the variable region of *N. crassa A* relative to the variable regions of *N. crassa a*, *N. sitophila A* and *N. sitophila a* is largely responsible for the apparent lesser level of sequence similarity between the mating types of *N. crassa A* and *a* strains.

A and *a* strains of the other three species have quite dissimilar variable regions. In all cases, a high level of sequence similarity begins at the border of the *A* and *a* idiomorphs at the exact nucleotide determined previously for the *N. crassa* idiomorphs (points C_A and C_a in Figures 4 and 5). In *N. intermedia* and *N. tetrasperma*, however, this intermating type identity is maintained only for 59 bp (the mating type common region); beyond this region, the variable regions of the *A* and *a* strains show only 21% similarity and so are essentially unrelated. *N. discreta* contains a similar mating type common region, although displaced because of the apparent 90-bp insertion between the *a* idiomorph and the *a* common region of *N. discreta a* (discussed above). The variable regions of *N. discreta A* and *a* show only 20% similarity to each other, *i.e.*, they are also unrelated. In each of these three latter species, the dissimilarity between the variable regions of each mating type extends for 3–5 kb (see APPENDIX). No mating type-specific islands of homology were present in the sequenced portion of these variable regions, although the Southern blot analysis summarized in Figure 1A suggests that a relatively large island of homology is common to the variable regions of the *A* and *a* strains of *N. discreta*. This region in the variable regions of *N. discreta A* and *a* is also present in the variable regions of each mating type in both *N. crassa* and *N. sitophila*. Beyond this large island of homology, the variable regions of the *A* and *a* strains of *N. discreta* continue to be dissimilar for ~3 kb.

Evolutionary relationships between Neurospora species: Inspection of the variable regions suggested that *N. crassa* and *N. sitophila* are more closely related to each other than to the other species and that *N. tetrasperma* and *N. intermedia* form a similar subgroup. Due to the absence of significant sequence similarity in the variable regions of different subgroups, it is not possible to construct a single meaningful phylogenetic tree using DNA sequence from this region. To see whether the phylogenetic relationships of the mating type idiomorphs correspond with the apparent relationships evident from the variable regions, we sequenced the *mt-A1* genes of previously uncharacterized Neurospora species and used these, along with the previously sequenced *mt-A1* genes, in a phylogenetic analysis.

Using two types of phylogenetic analyses, PAUP 3.0 (maximum parsimony) (SWOFFORD 1990) and the Clustal program (distance matrix, DNASTAR), we constructed phylogenetic trees of these species from the primary sequences of the *mt-A1* genes (Figure 6). In both analyses, *N. africana* and *N. discreta* are relatively distant from a more closely related group, which in-

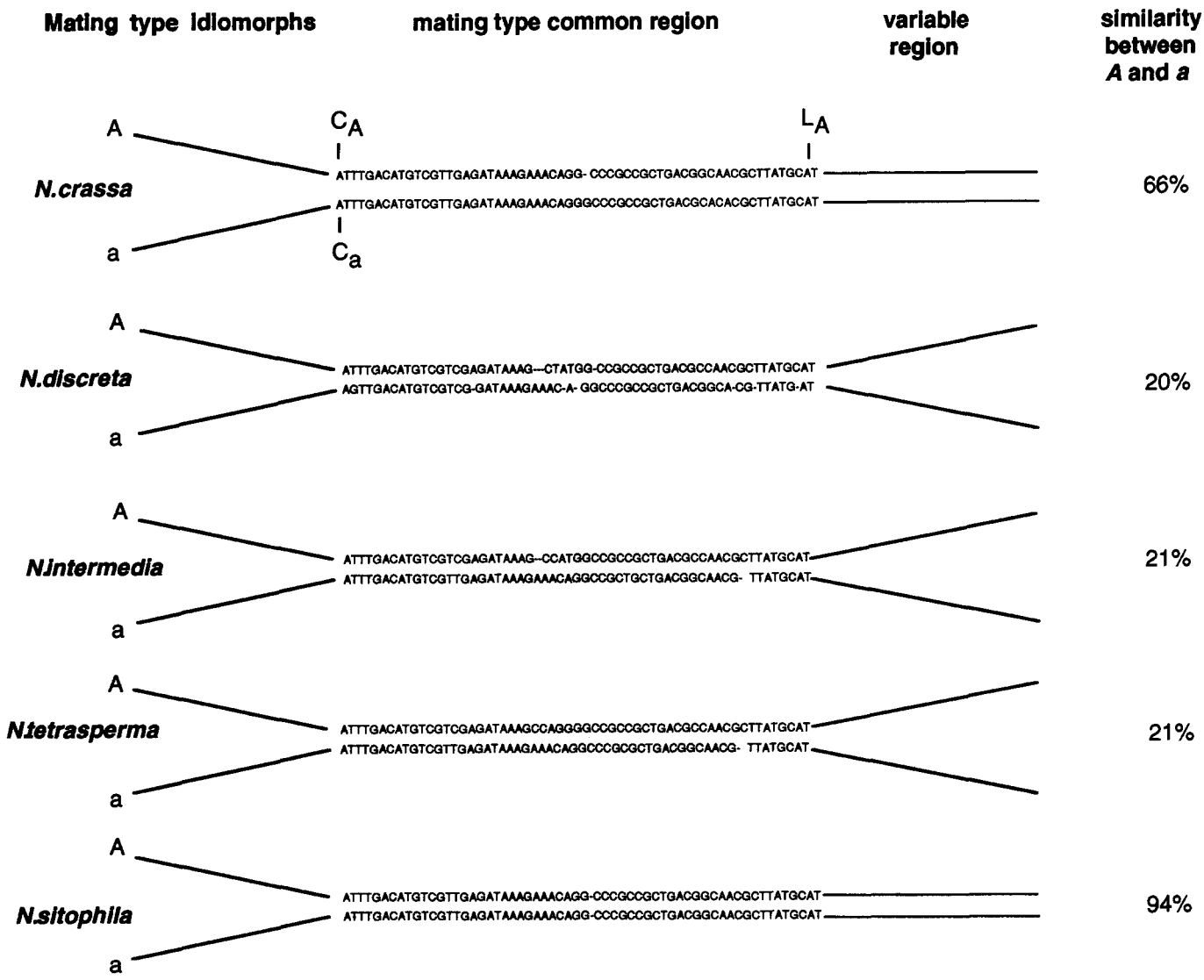

FIGURE 5.—Comparison of variable regions from *A* and *a* strains. The oblique nonparallel lines on the left indicate the position of known *A* and *a* idiomorphs. The sequence shown is the region of DNA flanking the idiomorphs that is identical in the *A* and *a* strains of each species (the mating type common region). Lines to the right indicate the positions of the variable regions while percent values next to each species pair indicate sequence similarity for the 800–1000 bp that has been sequenced from the variable regions; parallel, horizontal lines indicate similarity in sequence, while oblique, nonparallel lines indicate absence of similarity.

cludes the heterothallic *N. crassa, N. intermedia, N. sitophila* strains and *N. tetrasperma.* The relatedness of the *mt-A1* gene of *N. tetrasperma* to those of *N. crassa, N. intermedia,* and *N. sitophila* (also shown in the Clustal analysis; data not shown) suggests that this pseudohomothallic species arose after the divergence of this subgroup from its common ancestor with the heterothallic *N. discreta.*

As Table 5 indicates, all *mt-A1* genes show a very high level of sequence similarity. Thus, the abrupt shift from the well conserved idiomorphs to the very dissimilar variable regions suggests that an unusual mechanism is operating in this region of the genome to produce such sequence variation (see DISCUSSION). All of the variable region DNA sequences analyzed above were compared to the EMBL/Genbank database (as of December 1994) and no significant similarity to any sequence was found. If homologs of any of these sequences are present in other genera, they have not yet been analyzed by researchers working on those organisms.

DISCUSSION

We have shown that within the 3–5-kb region immediately adjacent to the centromere-proximal flank of the *A* and *a* mating type idiomorphs in the genus Neurospora there exists a region of great dissimilarity between species; we will call this the variable region. We have examined it both by DNA sequence, which allows us to set quantitative values of similarity or dissimilarity,

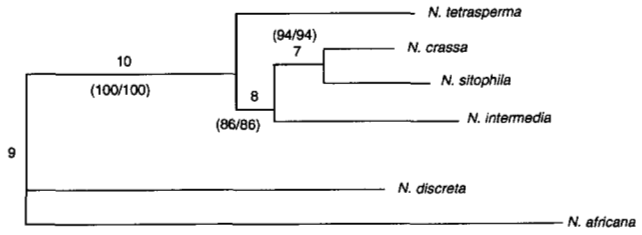

FIGURE 6.—Phylogenetic tree of *mt-A1* ORFs of all *A* mating types. The sequence of the *N. crassa mt-A1* gene is taken from GLASS *et al.* 1990a; the sequence of the *N. africana mt-A1* gene is taken from GLASS and SMITH 1994. The *mt-A1* genes from *N. intermedia, N. tetrasperma, N. discreta,* and *N. sitophila* were sequenced as described in MATERIALS AND METHODS. The *mt-A1* genes were initially aligned using the PILEUP option of the Wisconsin GCG program (DEVEREUX *et al.* 1984). The tree shown is the result of an exhaustive search with PAUP version 3.0 using the open reading frame with intron for each *mt-A1* gene. All characters were unordered and weighted equally. The numbers shown at each branch point are the confidence levels for bootstrapping of 500 replicates for each of two different methods (branch and bound/heuristic). The midpoint rooting option was used, but similar relationships and confidence levels were obtained when either the *N. africana, N. crassa,* or *N. tetrasperma mt-A1* genes were designated as the outgroup. The phylogram option was selected so that the branch lengths reflect the number of sequence differences (number not in parentheses above each branch).

and by Southern blotting, which allows us to rule out the possibility that complete dissimilarity of a flanking sequence can be explained, trivially, as being due to a chromosomal rearrangement. For the strains we have analyzed, which we take as being representative of each species, this region is in part species-specific; in some, but not all, of the species studied, the variable region is also extremely different between the two mating types of a particular species. No comparable differences between the species or mating types were seen in the centromere-distal flank of the *A* and *a* idiomorphs.

These variable regions form subgroupings according to species. These subgroupings are similar in strains of *A* and *a* mating type. The variable regions in the *A* mating types of *N. crassa* and *N. sitophila* are very similar to one another, as are the variable regions, of the *a* mating types. The variable regions in the *A* mating types of *N. intermedia* and *N. tetrasperma* are similar to one another, as are the variable regions of the *a* mating types. The *A* subgroups (*crassa/sitophila vs. intermedia/tetrasperma*) are unrelated to each other at the sequence level, as are the *a* subgroups. The sequenced portions of the *N. discreta* variable regions appear distinct by sequence analysis from that of both of the above subgroups.

There exist some less well conserved regions of similarity between species within the variable regions. Sequence analyses of the variable region from *N. africana, N. intermedia A,* and *N. discreta A* showed an island of homology of ~300–400 nucleotides, which are 50–60% similar to one another. Different additional islands

TABLE 5

Similarity of *mt-A1* ORFs from various Neurospora species

	Percent similarity					
	1	2	3	4	5	6
N. africana (1)	—					
N. crassa (2)	88	—				
N. discreta (3)	88	90	—			
N. intermedia (4)	89	98	90	—		
N. sitophila (5)	88	97	89	96	—	
N. tetrasperma (6)	89	98	89	97	96	—

of homology are present between individual pairs of variable regions. Islands of homology can occur in variable regions of both the *A* and *a* mating types. Other similarities outside the sequenced portions were identified by Southern analysis.

For the species *N. tetrasperma, N. intermedia,* and *N. discreta,* the sequenced portions of the variable regions of the *A* and *a* strains show only 20–21% sequence similarity, *i.e.,* the variable regions of the two mating types are unrelated. Southern analysis shows that this mating type difference extends for 3–5 kb. DNA hybridization analysis shows that similar variable regions are present in independently isolated strains of a given mating type of a particular species isolated from widely separated locations (*viz.,* in *N. tetrasperma, N. intermedia,* and *N. discreta*). This suggests a relatively recent common ancestor for strains within a species; this is, of course, consistent with the correlation of fertility with species designations. Assessment of the full extent of the similarities and differences will require more sequence analysis.

The left border marking the beginning of the variable regions in both mating types is well defined. In the *A* mating type of each species, there exists a region of 57–59 bp that is designated the *A* mating type common region. In the *a* mating type of each species, there exists a region of ~140 bp that is designated the *a* common region. The variable regions begin immediately beyond the mating type common region. The three species in which the variable regions of the *A* and *a* mating types are highly diverged within a species nevertheless share a common feature, the above mentioned mating type common region, comprising 57–59 bp of near identity both between mating types and between species. The striking conservation of the common region suggests that it has a functional role (such as a pairing site for homologous chromosomes during meiosis). Because the mating type common region is conserved even between *N. africana* and the heterothallic species (the greatest evolutionary distance), this region is probably not subject to the extensive sequence divergence that has occurred on its centromere-proximal side.

All of the true homothallic Neurospora species that

we have examined contain a variable region of ~7 kb. The size of this variable region appears similar in all of the homothallic species. Southern hybridization analysis shows the variable region of these homothallic species to be well conserved. This region is, however, very different at both the sequence level and by Southern hybridization from the corresponding variable regions of the heterothallic and pseudohomothallic Neurospora species. This could indicate that this region has a functional relation to the difference between homothallism and heterothallism in Neurospora.

We have found that the variable regions of *N. crassa A* and *a* give rise to a transcript (T. A. RANDALL and R. L. METZENBERG, unpublished results); this transcript is encoded within the *Acc*I-*Acc*I-*Eco*RI interval of Figure 1A. The size of the transcript from the *A* strain is different from that of the *a* strain, and the abundance of these transcripts is elevated under conditions favoring mating (nitrogen starvation). The *A*-specific transcript is regulated either directly or indirectly by the product of the *mt-A1* ORF because it is absent from a strain carrying a frameshift mutation within *mt-A1*. The data suggest an involvement of the products of these transcripts in either mating or sexual development. The second pair of transcripts described in the Introduction is immediately to the right of the variable regions in the area in which a high level of homology is found between all heterothallic and pseudohomothallic strains.

The subgroupings of the variable regions of the *A* and *a* strains noted above are consistent between mating types and with the tree we have constructed from the *mt-A1* genes. Both the hybridization and sequence data from the variable regions and the phylogenetic tree constructed for the *mt-A1* genes suggest a different evolutionary history from that proposed previously on the basis of RFLP analysis (NATVIG *et al.* 1987; TAYLOR and NATVIG 1989). These earlier studies (involving many more individual isolates of *N. crassa, N. intermedia, N. sitophila* and *N. tetrasperma*) suggested that the three heterothallic species are more closely related to each other than to the pseudohomothallic species. Our results also indicate that *N. tetrasperma* is the outgroup among these four species. However, our tree based on the *mt-A1* genes also suggests that *N. crassa* and *N. sitophila* are more closely related than previously thought and that *N. intermedia* and *N. tetrasperma* are closely related. The similarity of the variable regions of the *A* and *a* strains, in addition to the phylogenetic tree based on the *mt-A1* gene, indicates a very close relationship between *N. intermedia* and *N. tetrasperma* that had not been suggested by previous analyses. This could indicate that *N. tetrasperma* is descended from *N. intermedia* (or one of its recent ancestors). In previous analyses (TAYLOR and NATVIG 1989) many more *N. crassa* and *N. intermedia* isolates were examined, but the results did not indicate whether either species is monophyletic. A more definitive placement of these two species in relation to *N. sitophila* will require the sequence analysis of more isolates, and probably more genes from each isolate.

Both trees also suggest that the heterothallic *N. discreta* is more distantly related to the above strains than they are to one another; this suggests that *N. tetrasperma* evolved after the divergence of *N. discreta* from its common ancestor with the other heterothallic species. The relationship suggested between *N. discreta, N. crassa,* and *N. intermedia* is also supported by an evolutionary analysis of the well conserved *frq* gene; this places *N. discreta* relatively distant from these other two heterothallic species [this emerged from an analysis that did not address the placement of *N. tetrasperma* (MERROW and DUNLAP 1994)]. The true homothallic representative, *N. africana,* is most distant from the other species analyzed although the PAUP analysis does not distinguish whether *N. africana* or *N. discreta* is closer to the other strains; this determination may require further analysis of other homothallic Neurospora species. At first glance, the data suggest that the distinction between homothallism and heterothallism-pseudohomothallism is the most ancient point of divergence between these species. However, sexual isolation or absolute inbreeding is an ideal situation for rapid genetic drift, and it is possible that the "molecular clock" is speeded up so greatly in homothallic isolates that nucleotide substitutions are a poor indication of chronological time. This could be true for both *N. tetrasperma* and *N. africana.*

The abrupt change in sequence similarity at the idiomorph-proximal border of the variable regions suggests that an unusual evolutionary event has taken place in this region (more than random genetic drift). Two basic mechanisms could account for the presence of these variable regions. Horizontal transfer of viral or transposable elements to this region could have occurred repeatedly (KIDWELL 1993). However, comparison of the sequences obtained for each of these variable regions to the Genbank/EMBL library revealed no significant similarity to any known sequence. If each of these variable regions is the result of a horizontal transfer, they represent previously unidentified DNA elements. This hypothesis would also require several independent insertions of different DNA elements at identical points adjacent to the *A* and *a* idiomorphs or extreme divergence of a very ancient insertion, and both these alternatives seem implausible.

The sequenced regions of the variable regions are so different as to appear on first inspection to be unrelated to each other. If these sequences are ancestrally related, a very high rate of fixation of mutations would be needed to account for this level of difference. This elevated rate of mutation would also have to be limited to the variable regions themselves, because our sequencing and hybridization analyses indicate that DNA flanking both sides of the variable regions has diverged very

little. The sequence of these variable regions gives no clue as to their function. No major ORFs are seen within the limited region that has been sequenced, though many small ORFs (100–200 bp) have been noted in all the variable regions. A more complete sequence analysis of the entire 3–5-kb variable regions of each species along with gene disruptions may be needed to uncover the function(s) of this region.

Such sequence dissimilarity of regions at allelic positions between members of different species has not, to our knowledge been described. There are, however, several well characterized differences between members of the same species. Many of these are in fungal systems in which mating type idiomorphs have been described (see Introduction). In basidiomycetous fungi, several studies of mating-related genes have shown a high level of dissimilarity. In *Schizophyllum commune*, different Y and Z alleles at Aa loci show sequence identities of 42–54% (STANKIS *et al.* 1992). In *Coprinus cinereus*, different alleles within the Aa and Ab locus do not cross-hybridize at high stringency, although exact sequence comparisons are not given (KÜES *et al.* 1992). *Ustilago maydis* has two unlinked loci involved in mating and sexual development. The *a1* and *a2* alleles are very similar to idiomorphs in that they show no similarity at the DNA level (FROELIGER and LEONG 1991; BÖLKER *et al.* 1992). Each *b* locus contains a pair of homeodomain proteins. These proteins have constant and variable domains (KRONSTAD and LEONG 1990). Those variable regions that have been studied show amino acid identities of 44% (*bW* alleles) (GILLISSEN *et al.* 1992) and 60% (*bE* alleles) (SCHULZ *et al.* 1990). Plant systems with numerous incompatibility alleles also show significant variation (as low as 40% amino acid identities between different *S* alleles of various solanaceous species) (IOERGER *et al.* 1990). Two sequenced *S* alleles of *Petunia inflata* show sequence differences (74% and 84% identity for exons 1 and 2, respectively) (COLEMAN and KAO 1992) and others do not cross-hybridize at high stringency. Three alleles of the S locus of *Nicotiana alata* show hypervariable regions with 56% amino acid identity (ANDERSON *et al.* 1989).

Theories of speciation suggest that mechanisms must exist to allow increased variability in genes encoding determinants of speciation, although little experimental evidence has been found to support these theories (GOODENOUGH 1985). There are, however, systems in which an accelerated rate of sequence divergence has been seen to result in a change in function of related genes (HILL and HASTIE 1987) including one involving sequences in mammals adjacent to the sex determining locus SRY (TUCKER and LUNDRIGAN 1993; WHITFIELD *et al.* 1993).

This study raises several questions. How did the variable regions arise? The mechanism is not known; but it could be an extension of that which resulted in the evolution of the *A* and *a* idiomorphs themselves. How common is such a localized divergence between species within other genera? The variable region could represent a unique type of DNA element that could either be actively involved in the speciation process or it could be a relic of past speciation events. A systematic investigation of the extent of such differences within the genus Neurospora would seem warranted; additional variable regions could be present at other loci within this genus. Identification and characterization of any species-specific genes within the variable regions could aid in understanding species recognition and cell-cell interactions. In addition, the present study suggests that it may be possible to identify regions of the genome specific to either heterothallic or homothallic species. This may yield insight into the nature of these two common mating strategies found in eukaryotes as diverse as filamentous fungi and unicellular algae.

We thank MARY BERBEE, JAMES CROW, WILLIAM ENGELS, JEFF GROTELUESCHEN, DAVID JACOBSON, DON NATVIG, and WAYNE VERSAW for helpful discussion and criticism, and MARY ANN NELSON for her reworking of this manuscript beyond the call of collegiality. We are grateful for support from a U.S. Public Health Service grant, GM-08995, to R.L.M., and to the Merck Postdoctoral Fellowship Program for support of T.A.R. during much of this work.

LITERATURE CITED

ANDERSON, M. A., G. I. MCFADDEN, R. BERNATZKY, A. ATKINSON, T. ORPIN *et al.*, 1989 Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. Plant Cell **1:** 483–491.

BEATTY, N. P., M. L. SMITH and N. L. GLASS, 1994 Molecular characterization of mating-type loci in selected homothallic species of Neurospora, Gelasinospora and Anixiella. Mycol. Res. **98:** 1309–1316.

BÖLKER, M., M. URBAN and R. KAHMANN, 1992 The *a* mating-type locus of *Ustilago maydis* specifies cell signaling components. Cell **68:** 441–450.

COLEMAN, C. E., and T.-H. KAO, 1992 The flanking regions of two *Petunia inflata S* alleles are heterogeneous and contain repetitive sequences. Plant Mol. Biol. **18**: 725–737.

DEBUCHY, R., and E. COPPIN, 1992 The mating types of *Podospora anserina*: functional analysis and sequence of the fertilization domains. Mol. Gen. Genet. **233:** 113–121.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

FROELIGER, E. H., and S. A. LEONG, 1991 The *a* mating-type alleles of *Ustilago maydis* are idiomorphs. Gene **100**: 113–122.

GILLISSEN, B., J. BERGEMANN, C. SANDMANN, B. SCHROEER, M. BÖLKER, *et al.*, 1992 A two-component regulatory system for self/nonself recognition in *Ustilago maydis*. Cell **68:** 647–657.

GLASS, N. L., and G. A. KULDAU, 1992 Mating type and vegetative incompatibility in filamentous ascomycetes. Annu. Rev. Phytopathol. **30:** 201–224.

GLASS, N. L., and L. LEE, 1992 Isolation of *Neurospora crassa A* mating type mutants by repeat induced point (RIP) mutation. Genetics **132:** 125–133.

GLASS, N. L., and M. L. SMITH, 1994 Structure and function of a mating-type gene from the homothallic species *Neurospora africana*. Mol. Gen. Genet. **244:** 401–409.

GLASS, N. L., J. GROTELUESCHEN and R. L. METZENBERG, 1990a *Neurospora crassa A* mating-type region. Proc. Natl. Acad. Sci. USA **87:** 4912–4916.

GLASS, N. L., R. L. METZENBERG and N. B. RAJU, 1990b Homothallic Sordariaceae from nature: the absence of strains containing only the *a* mating type sequence. Exp. Mycol. **14:** 274–289.

GOODENOUGH, U. W., 1985 An essay on the origins and evolution of eukaryotic sex, pp. 123–140 in *The Origins and Evolution of*

Sex, edited by H. O. HALVORSON and A. MONROY. Alan R. Liss, New York.

GRIFFITHS, A. J. F., 1982 Null mutants of the A and a mating type alleles of *Neurospora crassa*. Can. J. Genet. Cytol. **24:** 167–176.

HILL, R. E., and N. D. HASTIE, 1987 Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature **326:** 96–99.

IOERGER, T. R., A. G. CLARK and T.-H. KAO, 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA **87:** 9732–9735.

JACOBSON, D. J., 1992 Control of mating type heterokaryon incompatibility by the *tol* gene in *Neurospora crassa* and *N. tetrasperma*. Genome **35:** 347–353.

KIDWELL, M. G., 1993 Lateral transfer in natural populations of eukaryotes. Annu. Rev. Genet. **27:** 235–256.

KRONSTAD, J. W., and S. A. LEONG, 1990 The *b* mating-type locus of *Ustilago maydis* contains variable and constant regions. Genes Dev. **4:** 1384–1395.

KÜES, U., W. V. J. RICHARDSON, A. M. TYMON, E. S. MUTASA, B. GÖTTGENS *et al.*, 1992 The combination of dissimilar alleles of the Aa and Ab gene complexes, whose proteins contain homeo domain motifs, determines sexual development in the mushroom *Coprinus cinereus*. Genes Devel. **6:** 568–577.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MERROW, M. W., and J. C. DUNLAP, 1994 Intergeneric complementation of a circadian rhythmicity defect: phylogenetic conservation of structure and function of the clock gene *frequency*. EMBO J. **13:** 2257–2266.

METZENBERG, R. L., and S. K. AHLGREN, 1973 Behaviour of *Neurospora tetrasperma* mating-type genes introgressed into *N. crassa*. Can. J. Genet. Cytol. **15:** 571–576.

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.

PERKINS, D. D., 1977 Behavior of *Neurospora sitophila* mating-type alleles in heterozygous duplications after introgression into *Neurospora crassa*. Exp. Mycol. **1:** 166–172.

PERKINS, D. D., and N. B. RAJU, 1986 *Neurospora discreta*, a new heterothallic species defined by its crossing behavior. Exp. Mycol. **10:** 323–338.

PERKINS, D. D., B. C. TURNER and E. G. BARRY, 1976 Strains of Neurospora collected from nature. Evolution **30:** 281–313.

RAJU, N. B., 1977 Meiotic nuclear behaviour and ascospore formation in five homothallic species of Neurospora. Can. J. Bot. **56:** 754–763.

RAJU, N. B., 1992 Functional heterothallism resulting from homokaryotic conidia and ascospores in *Neurospora tetrasperma*. Mycol. Res. **96:** 103–116.

SCHULZ, B., F. BANUETT, M. DAHL, R. SCHLESINGER, W. SCHÄFER *et al.*, 1990 The *b* alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. Cell **60:** 295–306.

SINCLAIR, A. H., P. BERTA, M. S. PALMER, J. R. HAWKINS, B. L. GRIFFITHS *et al.*, 1990 A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature **346:** 240–244.

STANKIS, M. M., C. A. SPECHT, H. YANG, L. GIASSON, R. C. ULLRICH *et al.*, 1992 The Aa mating locus of *Schizophyllum commune* encodes two dissimilar multiallelic homeodomain proteins. Proc. Natl. Acad. Sci. USA **89:** 7169–7173.

STABEN, C., and C. YANOFSKY, 1990 *Neurospora crassa a* mating-type region. Proc. Natl. Acad. Sci. USA **87:** 4917–4921.

STEVENS, J. N., and R. L. METZENBERG, 1982 Preparing Neurospora DNA: some improvements. Neurospora Newsl. **29:** 27–28.

SWOFFORD, D., 1990 PAUP: phylogenetic analysis using parsimony. A computer program distributed by the Illinois Natural History Survey, Champaign, IL.

TAYLOR, J. W., and D. O. NATVIG, 1989 Mitochondrial DNA and evolution of heterothallic and pseudohomothallic Neurospora species. Mycol. Res. **93:** 257–272.

TUCKER, P. K., and B. L. LUNDRIGAN, 1993 Rapid evolution of the sex determining locus in old world mice and rats. Nature **364:** 715–717.

TURGEON, B. G., H. BOHLMANN, L. M. CIUFFETTI, S. K. CHRISTIANSEN, G. YANG *et al.*, 1993 Cloning and analysis of the mating type genes from *Cochliobolus heterostrophus*. Mol. Gen. Genet. **238:** 270–284.

WHITFIELD, L. S., R. LOVELL-BADGE and P. N. GOODFELLOW, 1993 Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature **364:** 713–715.

Communicating editor: R. H. DAVIS

APPENDIX

Sequence comparison of the variable regions from *A* strain

Similarity between the variable regions of *N. crassa A* and *N. sitophila A:* These variable regions have 95% similarity between bp 1–272. From bp 272–432 there is a lower level of similarity (44%), and between *N. crassa* (bp 432–632) and *N. sitophila* (bp 632–842) the sequence similarity is 83%. The variable region of *N. crassa A* lacks a 200-bp sequence that is present in the variable regions of *N. sitophila A* (bp 433–631), *N. sitophila a*, and *N. crassa a* (see below). The gap penalty because of this deletion accounts for the seemingly modest level of similarity between these regions, which is otherwise very high. Because sequence corresponding to this region of the variable region of *N. sitophila A* is also present in the variable regions of *N. sitophila a* and *N. crassa a*, we consider it likely that a deletion occurred in the variable region of *N. crassa A* rather than an insertion in the other three strains.

Similarity between the variable regions of *N. intermedia A* and *N. tetrasperma A*: The sequenced portions of the variable regions of *N. intermedia A* and *N. tetrasperma A* (designated the *N. intermedia*-like subgroup) have an overall sequence similarity of 47%, but the extent of similarity varies across the sequenced part of the variable region. Between the variable region of *N. intermedia A* (bp 1–460) and the variable region of *N. tetrasperma A* (bp 1–483) the sequence similarity is 74%, but in the remaining sequenced region the similarity drops off to 21% (*i.e.*, they are unrelated). The variable regions of the *N. crassa* and *N. intermedia* subgroups have little sequence similarity.

Sequence comparisons between variable regions of the *A* strains of other species: The variable regions of *N. africana* and *N. discreta*, which have extensive similarity, have no significant regions of similarity to the variable regions of *N. crassa* and *N. sitophila* in the 1 kb that we have sequenced; see Table 3. A region in the variable region of *N. africana* (bp 148–700) has 59% similarity to the variable region of *N. discreta A* (bp 1–602). The variable region of *N. africana* (bp 148–608) has 66% similarity to the variable region of *N. intermedia A* (bp 199–652) and from bp 150–265 has 64% similarity to the variable region of *N. tetrasperma A* (bp 199–358).

Similarity between variable regions of *N. intermedia a*, *N. tetrasperma a*, and *N. discreta a*

The only major departure from the high degree of sequence similarity between the variable regions of *N.*

intermedia and *N. tetrasperma a* is an apparent insertion in the variable region of *N. intermedia a* from bp 305–397 (or a deletion from *N. tetrasperma*). In addition, the variable region of *N. discreta a* (bp 137–285) is 68% similar to the variable regions of *N. intermedia a* and *N. tetrasperma a* (bp 82–247 in each). There is 72% similarity between the variable region of *N. discreta a* (bp 292–414) and the variable region of *N. tetrasperma a* (bp 445–561).

Approximate positions of the right borders of the variable regions

The two mating types of *N. intermedia* return to similarity within the *Eco*RI-*Hind*III fragment of pniF2 (Figure 3A; see below). Beyond this point, both *N. intermedia* mating types also show homology with *N. crassa*, and to all other *Neurospora* species analyzed (data not shown). A return to homology with *N. crassa* beyond the *Cla*I-Cl site also occurs with the variable region of *N. tetrasperma A*; however, DNA from *N. tetrasperma A* does not hybridize to the *N. crassa*-derived probe 6 of Figure 1A. This is represented in the third column of floating boxes by a box unique to *N. tetrasperma A*. Therefore, the point at which this strain returns to homology to *N. crassa* is defined as the right border of the variable region for all species.

Hybridization analysis using the cloned region of the variable region of *N. tetrasperma* A (Figure 3A; see below) indicates that the mating type-specific portion of the variable region of *N. tetrasperma A* has a size of ~5 kb (a return to homology with *N. tetrasperma a* occurs beyond the DNA corresponding to the cloned plasmid pntAF shown in Figure 3A). As shown in Figure 1A, major portions of the variable regions of *N. intermedia a* and of *N. tetrasperma a* have homology to *N. crassa A* and *a*. Approximately 2 kb of the variable regions of the *a* strains of these two species shows no significant DNA homology to the corresponding *A* strains. Southern blot analysis using probes from the variable region of *N. discreta A* did not detect hybridization to *N. discreta a* (the cloned insert of pndCla2 and part of pndJGA [discussed below]), indicating that ~5 kb of the variable regions of *N. discreta A* and *a* are mating type-specific. Like the variable regions of the *a* mating types of *N. intermedia* and *N. tetrasperma*, the variable region of *N. discreta a* shows more regions with homology to both mating types of *N. crassa* than does the variable region of *N. discreta A*.

Sequence comparison between the centromere-proximal region of the *A* idiomorphs of all species

Two distinctive features were revealed within the extreme centromere-proximal region of the *A* idiomorphs themselves. The first feature is an 81-bp insertion in the *N. crassa* idiomorph that is not present in any other *A* idiomorphs (Figure 4A). PCR analysis using one primer whose sequence is conserved within the *mt-A1* gene of all species and a second primer within the 81-bp *N. crassa A* insertion amplified an appropriately-sized DNA product in other *N. crassa A* strains in our collection, suggesting that all contain this 81-bp region. No DNA fragment was amplified with these primers using genomic DNA from any strains of *N. tetrasperma A*, *N. intermedia A*, or *N. sitophila A*. This suggests that this region is absent in these isolates (as was the case in the sequenced representative of these three species). The second feature, a 4-bp insertion in *N. crassa A* 9 bp downstream of the 81-bp insertion, was not present in any of the sequenced representatives of the other species. Two strains of *N. crassa A* were sequenced over this region (Oak Ridge wild type and Mauriceville-1c), and both had the 81-bp and 4-bp insertions. Neither of two *N. sitophila A* strains sequenced had these insertions. The presence of this 4-bp insertion was not examined in other representatives of the studied species. As we have not been able to identify any major open reading frames or transcripts in this region, and no systematic mutagenesis of this region of the *A* idiomorph has been performed, the significance of these differences is unknown.

Species-specific differences in the variable regions

The data presented above indicate that the variable regions of the *A* and *a* mating types can be mating type-specific, as well as species-specific. Further DNA hybridization analysis with selected subfragments from the clones shown in Figure 3A showed that some DNA fragments within the variable region of *N. intermedia A* that are homologous to *N. tetrasperma A* are not present within the genome of any other species. These results are summarized in Figure 1A. DNA fragment B of Figure 3A showed homology only to the variable regions of *N. intermedia A* and *N. tetrasperma A*. A sequence from the variable region of *N. discreta A* contiguous to the *A* common region showed no similarity to any other variable regions. However, a more centromere-proximal region of DNA within the variable region showed homology to sequences in the *A* mating types of two of the other species.