Multiple Origins and nrDNA Internal Transcribed Spacer Homeologue Evolution in the *Glycine tomentella* (Leguminosae) Allopolyploid Complex

Jason T. Rauscher,*[,1] Jeff J. Doyle* and A. H. D. Brown[†]

**L. H. Bailey Hortorium, Cornell University, Ithaca, New York 14853 and [†]Centre for Plant Biodiversity Research, Commonwealth Scientific and Industrial Research Organization Plant Industry, Canberra, ACT 2601, Australia*

Manuscript received March 26, 2003
Accepted for publication November 2, 2003

ABSTRACT

Despite the importance of polyploidy in the evolution of plants, patterns of molecular evolution and genomic interactions following polyploidy are not well understood. Nuclear ribosomal DNA is particularly complex with respect to these genomic interactions. The composition of nrDNA tandem arrays is influenced by intra- and interlocus concerted evolution and their expression is characterized by patterns such as nucleolar dominance. To understand these complex interactions it is important to study them in diverse natural polyploid systems. In this study we use direct sequencing to isolate and characterize nrDNA internal transcribed spacer (ITS) homeologues from multiple accessions of six different races in the *Glycine tomentella* allopolyploid complex. The results indicate that in most allopolyploid accessions both homeologous nrDNA repeats are present, but that there are significant biases in copy number toward one homeologue, possibly resulting from interlocus concerted evolution. The predominant homeologue often differs between races and between accessions within a race. A phylogenetic analysis of ITS sequences provides evidence for multiple origins in several of the polyploid races. This evidence for diverse patterns of nrDNA molecular evolution and multiple origins of polyploid races will provide a useful system for future studies of natural variation in patterns of nrDNA expression.

POLYPLOIDY is commonly recognized as having played a critical role in the evolution of plants, especially with respect to the origin of new plant species and higher taxa (Leitch and Bennett 1997; Bretagnolle *et al.* 1998; Soltis and Soltis 1999). The genomic consequences of polyploidy, especially allopolyploidy, however, are not well understood (Song *et al.* 1995; Osborn *et al.* 2003). While one of the presumed advantages of allopolyploidy is the genetic diversity and redundancy associated with fixed heterozygosity of homeologous loci, the fates of these loci following polyploidy can be complex and involve such phenomena as epigenetic regulation of gene expression (*e.g.*, Lee and Chen 2001) or the complete loss of a homeologous locus (see reviews in Soltis and Soltis 1999; Wendel 2000; Pikaard 2001).

This complexity is compounded for nuclear ribosomal DNA (nrDNA), since hundreds or thousands of copies of the genes are organized in tandem repeats at a locus and are subject to molecular evolutionary forces such as concerted evolution (Hamby and Zimmer 1992; Baldwin *et al.* 1995). For this reason, the fate of nrDNA loci after polyploidy is not well understood, and empirical studies have revealed a range of patterns, from the maintenance of both homeologous loci (*e.g.*, Sang *et al.* 1995) to the rapid loss of a locus or interlocus homogenization between homeologous loci (*e.g.*, Wendel *et al.* 1995).

Variation in ribosomal repeat copy number and the potential for concerted evolution between homeologous loci in polyploids may have important implications for understanding the control of ribosomal DNA expression. Nucleolar dominance is the preferential expression of the nucleolar organizing region of one progenitor species in an interspecific hybrid and results in a pattern in which expression of only one of the two nrDNA homeologues is observed in an allopolyploid (Pikaard 2000). It has been described in a few model systems (*e.g.*, Chen *et al.* 1998; Lewis and Pikaard 2001), but has yet to be studied thoroughly in a diverse natural polyploid complex. Describing the fate of homeologous nrDNA repeats in natural allopolyploids and identifying polyploid lineages with multiple origins is a necessary first step to provide a genetic and evolutionary context for understanding variation in patterns of nrDNA gene expression.

Nuclear ribosomal DNA contains regions such as the internal transcribed spacers (ITSs) that are highly variable and rich with phylogenetic information (Baldwin *et al.* 1995) and can therefore serve as a valuable tool, both in elucidating polyploid origins and in understanding the genomic consequences of polyploidy. Few studies, however, have characterized nrDNA ITS variation

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY433834–AY433934.

[1]*Corresponding author:* Departamento de Ciencias Biológicas, Apartado Aéreo 4976, Universidad de los Andes, Bogotá, D.C., Colombia. E-mail: rauscher@evoandes.net

and followed the fates of homeologous nrDNA loci in multiple accessions of a well-studied polyploid complex. One group that is particularly amenable to such a study is the *Glycine tomentella* polyploid complex. This complex is among the wild perennial relatives of the soybean, *G. max*, and has been collected extensively throughout its native range in Australia and the surrounding Pacific islands. Many accessions of this complex have also been characterized for reproductive isolation (Doyle *et al.* 1986), cytogenetics (Kollipara *et al.* 1994), isozymes (Doyle and Brown 1985; Singh *et al.* 1987; Kollipara *et al.* 1994), and most recently, DNA sequence variation at the histone H3D (Doyle *et al.* 2002) and nrDNA ITS loci (Kollipara *et al.* 1997; Hsing *et al.* 2001).

From these studies, it is known that the species *G. tomentella* is in fact a large polyphyletic complex of diploid ($2n = 38, 40$) and tetraploid ($2n = 78, 80$) cytotypes. The diploid "races" that have been identified within the complex are designated by crossing groups and include D1/D2, D3, D4, D5A, and D5B. Each is genetically distinct, represents a unique phylogenetic origin, and despite the lack of formal taxonomic recognition, should be considered different species (Kollipara *et al.* 1997; Brown *et al.* 2002). Allopolyploid lineages have arisen by crosses among these diploid races and with other species. Currently, six major tetraploid races are recognized and their diploid progenitors have been identified (Doyle *et al.* 2002). These include T1, formed originally by a cross between D1 and D3 (D1 × D3), T2 (D3 × D4), T3 (D3 × D5A), T4 (D3 × D5B), T5 (D1 × *G. clandestina*), and T6 (D1 × D5B). Evidence from a single-copy nuclear gene, histone H3-D, suggests that several of these polyploid taxa have originated more than once (Doyle *et al.* 2002), a phenomenon known to be common in the evolution of polyploids (*e.g.*, Cook *et al.* 1998; Segraves *et al.* 1999; Soltis and Soltis 1999).

In this study, we use direct sequencing of the nrDNA ITS region, using universal and repeat-specific primers, to characterize multiple accessions of six races of polyploids in the *G. tomentella* allopolyploid complex and assess qualitative differences in nrDNA copy number between homeologous loci. Phylogenetic analysis is then used to access the relationships between sequences from polyploid homeologous loci and loci from extant diploids to look for evidence of multiple origins within each polyploid race. In this way we shed light both on patterns of nrDNA evolution and on the origins and evolution of this polyploid complex.

MATERIALS AND METHODS

Sampling and DNA isolation: For most of the *G. tomentella* polyploid races, multiple accessions were obtained from the Commonwealth Scientific and Industrial Research Organization (CSIRO) native Australian Perennial Glycine collection. Locality data and voucher information for each accession are available in prior publications (Doyle *et al.* 1986, 2002) and at the Australian Plant Genetic Resources Information System website (http://www.dpi.qld.gov.au/extra/asp/AusPGRIS; the prefix "G" in accessions referred to in this article has been replaced by the prefix "AusIRC" in this database). In addition, several accessions of Glycine diploids were sampled to supplement ITS sequences from earlier data sets (Kollipara *et al.* 1997); specific accessions were chosen on the basis of a previous study (Doyle *et al.* 2002) that identified potential close relatives of polyploid sequences, as well as phylogenetically or geographically divergent taxa or populations. Chromosome counts for most accessions included in this study are available (see Doyle *et al.* 1986, 2002), and ploidy levels for all accessions have been confirmed by isozyme (Doyle and Brown 1985; Doyle *et al.* 1986) and histone H3D data (Doyle *et al.* 2002). Genomic DNA was isolated for previous studies of these accessions (Doyle and Brown 1989; Doyle *et al.* 2002).

Amplification and sequencing: The entire nrDNA ITS region was amplified for both polyploid and diploid accessions using primers ITS-5 Ang (Vazquez 2001; Rauscher *et al.* 2002) and ITS-4 (White *et al.* 1990); PCR conditions are described in detail elsewhere (Rauscher *et al.* 2002). The PCR buffer contained 20% DMSO to promote amplification of multiple ITS repeats (Buckler *et al.* 1997). Templates were purified and directly sequenced using the reverse (ITS-4) primer at the Cornell University Sequencing Facility on an ABI 3700 DNA sequencer. Several accessions of each polyploid race were also sequenced in the forward direction for confirmation. Resulting electropherograms were analyzed in Sequencher 4.1.2 and examined by eye for nucleotide additivity (double peaks) that would suggest the presence of more than one repeat type. Polymorphic sites were recorded as degenerate nucleotides using standard International Union of Biochemistry codes.

In polyploid accessions that produced clean (*i.e.*, monomorphic) sequence with universal ITS primers, repeat-specific and exclusion primers (Rauscher *et al.* 2002) were used in an attempt to amplify the second homeologous repeat. In addition, repeat-specific primers ITS-D1/D2 and ITS-cla were used to isolate the individual ITS repeats in the T5 polyploids (see Rauscher *et al.* 2002 for details). All specific and exclusion primers were paired with either the ITS-4 reverse primer or an alternative reverse primer (ITS-mr2: 5′-GTA GCC CCG CCT GAC CTG-3′) with a higher annealing temperature. Products of amplification with specific and exclusion primers were sequenced with the reverse primer and analyzed as described above. When amplification using specific or exclusion primers was weak, 100- to 200-μl reaction volumes were used to produce sufficient PCR product for direct sequencing. All sequences were exported to the program Se-Al v2.0a11 (Rambaut 1995) and manually aligned.

Phylogenetic analysis: For sequences that showed evidence of more than one ITS repeat type within a single homeologue, as shown by double peaks on the electropherogram, the two contributing sequences were inferred by comparison to monomorphic ITS sequences from the same data. In most cases, there was only a single polymorphic nucleotide site, making the inference of divergent sequences unambiguous. In the few accessions with multiple polymorphic sites, the signal of one repeat sequence was significantly stronger than the other and tended to be identical or nearly identical to other sampled sequences, facilitating the resolution of the two repeats without cloning.

Identical sequences were identified using MacClade 3.0 (Maddison and Maddison 1992) and all but a single representative of each unique sequence were removed prior to phylogenetic analysis. All unique, monomorphic sequences from diploid and polyploid accessions were combined into a

single data set and aligned manually using Se-Al v2.0 (Rambaut 1995). A maximum parsimony analysis was performed with PAUP* 4.0 (Swofford 2000) using a heuristic search with 20 random taxon addition sequences, TBR branch swapping, and gaps coded as missing data. Maxtrees was set at 20,000 and a strict consensus tree was constructed. The same analysis was performed for sequences from diploid accessions alone. Branch support was estimated for both the entire data set and the diploid data set using a bootstrap analysis consisting of 500 replicate heuristic searches with a single repetition of a random addition sequence and Maxtrees set to 100 for each replicate.

On the basis of the parsimony analysis of the entire data set, clades representing each of the major diploid races (D1/D2, D3, D4/*G. clandestina*, D5A, and D5B) were identified and separate data sets were constructed for each clade. Insertions and deletions were coded as binary characters and included in the analysis. Unrooted haplotype networks were used for each major clade containing both diploid and polyploid repeats because they clearly illustrate the variation, allow ambiguities to be highlighted, and may be more accurate than parsimony reconstructions when few molecular characters are available (Crandall 1994). These networks were constructed manually and then verified with the program TCS 1.13 (Clement *et al.* 2000; using the network-building algorithm of Templeton *et al.* 1992) and either heuristic or branch and bound searches in PAUP* 4.0 to make sure that all of the most parsimonious arrangements of the network had been considered. Due to the slightly downstream position of specific and exclusion primers, there was often an ~30- to 100-bp region of missing data on the 5′ end of sequences obtained from these primers. These missing data caused problems with the TCS analysis, so sequences that were identical to a diploid or polyploid accession at all other nucleotide positions were considered identical at the missing positions and removed from the analysis. In cases where the remaining sequence was not identical to any sampled diploid or polyploid sequences, the missing positions were scored as identical to the accession that was most similar at all other sites. Because these ambiguous characters affected branch lengths but not topological relationships, the decision was made to leave them in the analysis and label them on the resulting networks. Finally, possible rootings of each network were identified using the parsimony analysis described above.

RESULTS

Direct sequencing with universal ITS primers: When universal primers were used to amplify and then directly sequence the ITS region in *G. tomentella* polyploids, clean sequence of just a single homeologous parental ITS repeat type was found in all accessions of the T1, T2, T3, and T6 polyploids (Table 1). In two of these races, however, not all accessions resulted in the same parental homeologue being amplified. Of the 10 T1 (D1 × D3) plants sampled, 9 revealed the D3 homeologue, while 1 (G1392) gave the D1 homeologue. Similarly, among 10 T3 (D3 × D5A) polyploids sampled, 9 sequences were D3 homeologues and 1 (G1397) was the D5A homeologue. For all six accessions of T2 (D3 × D4), direct sequencing gave the D4 homeologue; the single accession of T6 (D1 × D5B) resulted in the D5B homeologue.

In the remaining polyploid races, T4 and T5, additive signal from the two parental homeologous ITS copies was visible in universal direct sequences from at least some accessions (Table 1). As previously reported, all three T5 (D1 × *G. clandestina*) accessions showed significant secondary peaks (Rauscher *et al.* 2002). Of the 12 T4 (D3 × D5B) accessions sampled, 6 resulted in clean sequence, 4 with the D3 homeologue, and 2 with the D5B homeologue. The remaining 6 accessions produced patterns of additive sequence consistent with the presence of both homeologous ITS copies. In the case of G1469, the signal from the D5B homeologue was very weak with respect to the dominant D3 signal, but visible upon close inspection of the electropherogram.

In one T4 polyploid accession, G1469, direct sequencing with universal ITS primers also resulted in two nucleotide positions with double peaks on the electropherogram that could not be accounted for by overlapping signal from the two homeologous repeats (D3 and D5B). Subsequent sequencing with specific primers (see below) showed these to be the result of two minor ITS repeats, one that was related to the D3 homeologue and the other to the D5B homeologue. In addition, one diploid accession (G2058, a D5A *G. tomentella*) and one tetraploid (G1956) that is not part of the core *G. tomentella* allopolyploid complex considered here had detectable minor repeats.

Specific-primer amplification and sequencing: In most accessions that gave clean sequence for a single homeologue with universal primers, the second ITS homeologue was readily amplified using repeat-specific primers. In T1 polyploids with a dominant D3 repeat, the D1 repeat amplified only weakly with the D1/D2-specific primer, but there was sufficient product for direct sequencing to confirm its identity. In the single T1 accession with a dominant D1 homeologue (G1392), the D3 repeat amplified strongly with the D3-specific primer. All T2 polyploids produced strong amplification with the D3-specific primer, as did the T6 polyploid with the D1/D2-specific primer.

Only a few accessions failed to yield a second ITS homeologue when amplified with specific primers. This included a single T4 accession (G2437) in which the D3 homeologue could not be amplified and six T3 polyploid plants. Among these T3 accessions, five (G1359, 1394, 1930, 2059, and 1766) had no amplifiable D5A homeologue and a single accession (G1397) had no D3 homeologue. In the case of the five accessions with no D5A, it is unlikely that any sequence other than the D3 homeologue was present since the exclusion primer ITS-nonD3 was used and resulted in no amplification.

Using specific primers, several minor variants of one or the other homeologous ITS repeats were detected in polyploid accessions. Most of these minor repeats, including a D3 repeat in G1392 (T1) and D5B repeats in four T4 accessions (G2468, G2470, G2476, and G1469) differed by just a single base pair from the major repeat sequence. In several T1 accessions and the single T6

TABLE 1

Results of direct sequencing with universal ITS primers

Polyploid race	Clean direct sequences	Additive direct sequences
T1 (D1 × D3)	D3: G1133, 1136, 1180,[a] 1274, 1288,[a] 1361, 1367, 1427, 1763[a]; D1: G1392	—
T2 (D3 × D4)	D4: G1134,[a] 1188,[a] 1286, 1393,[a] 1811, 1854	—
T3 (D3 × D5A)	D3: G1359, 1394, 1930, 1766, 2059, 2098, 2099, 2100, 2539; D5A: G1397	—
T4 (D3 × D5B)	D3: G1304, 1929, 1348, 3137; D5B: G2437, 2469	D3 + D5B: G1469, 1747, 2476, 2557, 2468, 2470
T5 (D1 × *G. clandestina*)	—	D1 + *G. clandestina*: G1487,[a] 1739,[a] 1969[a]
T6 (D1 × D5A)	D5B: G1945	—

The first column lists accessions that had clean sequence (no significant secondary peaks on the electropherogram) and identifies which of the two parental homeologues resulted. The second column shows accessions that resulted in an additive signal from both homeologues (double peaks or unreadable sequence on the electropherogram).

[a] Sequences were presented in a previous publication (Rauscher *et al.* 2002).

accession, minor D1 repeats that differed by 4–6 bp from the major D1 repeat were found. In each of these accessions the minor repeat could be inferred from the direct sequence because the signal of the minor D1 repeat was significantly weaker than that of the major repeat.

Phylogenetic analysis: A total of 104 accessions (62 diploids and 42 polyploids) were included in this study, including 37 for which sequences had been published previously (Kollipara *et al.* 1997; Hsing *et al.* 2001; Rauscher *et al.* 2002). From these, 155 ITS repeat sequences were obtained, 64 from diploids (62 major and 2 minor repeats) and 91 from polyploids (77 major and 14 minor repeats).

For the maximum-parsimony analysis of the entire data set there were 70 unique sequences. The heuristic search, limited to 20,000 trees, resulted in a set of trees with 285 steps, a consistency index of 0.74, and a retention index of 0.88. When only diploid accessions were included in the analysis the data set contained 48 unique sequences and resulted in the 56 shortest trees with a length of 268 steps, a consistency index of 0.75, and a retention index of 0.84. The strict consensus tree of the diploid analysis is shown in Figure 1 with bootstrap values (for both the diploid and the entire data sets) labeled on the major branches. Polyploid accessions from which ITS repeat sequences were obtained are shown in boxes to the side, with lines showing their affinities to diploid clades (specific relationships are shown in subsequent network figures). As shown in Figure 1, nearly all of the homeologous ITS sequences from polyploid accessions grouped with the diploid parental clades expected from previous studies (*e.g.*, Doyle *et al.* 2002). The only exception was accession G3137, a T4 (D3 × D5B) polyploid, in which the ITS-nonD3 exclusion primer amplified a D5A homeologue instead of the expected D5B.

For the network analysis, six major clades were identified on the basis of the original maximum-parsimony analysis (Figure 1), and a separate network was constructed for each (Figures 2–6). In the case of the D5A clade, the monophyly of the two subclades (D5A.1 and D5A.2) had bootstrap support of <50% (Figure 1). However, the presence of this clade in the strict consensus and support for D5A monophyly from other data sets (*e.g.*, Doyle *et al.* 2002) justified its analysis as a single network.

In the D1/D2 clade (Figure 2) there was little sequence variation among sampled diploids; three D1 accessions were identical and differed by two changes from the two D2 accessions. Three of the polyploid races (T1, T5, and T6) contained a D1 homeologue (Figure 1). In all of the T1 accessions the major repeat was identical to the diploid D1 sequence. Minor polyploid D1 repeat sequences had four to six nucleotide differences and branched from the node connecting the two diploid sequences (Figure 2; major repeats are labeled "a" and minor repeats "b"). Two of the accessions that had no detectable minor repeat (G1133 and G1274) were from the same geographical location (Brampton Island, Queensland). The third was G1392 (from Petford, Queensland), the only T1 in which the D1 homeologue was the dominant sequence using universal ITS primers. An accession from the only known population of the T6 race showed a similar pattern of polymorphism for the D1 homeologue. The major repeat differed by one nucleotide from the diploid D1 sequence; the inferred minor repeat differed by four nucleotides from the first (Figure 2). Finally, the D1 homeologue from the three T5 polyploids differed from the common

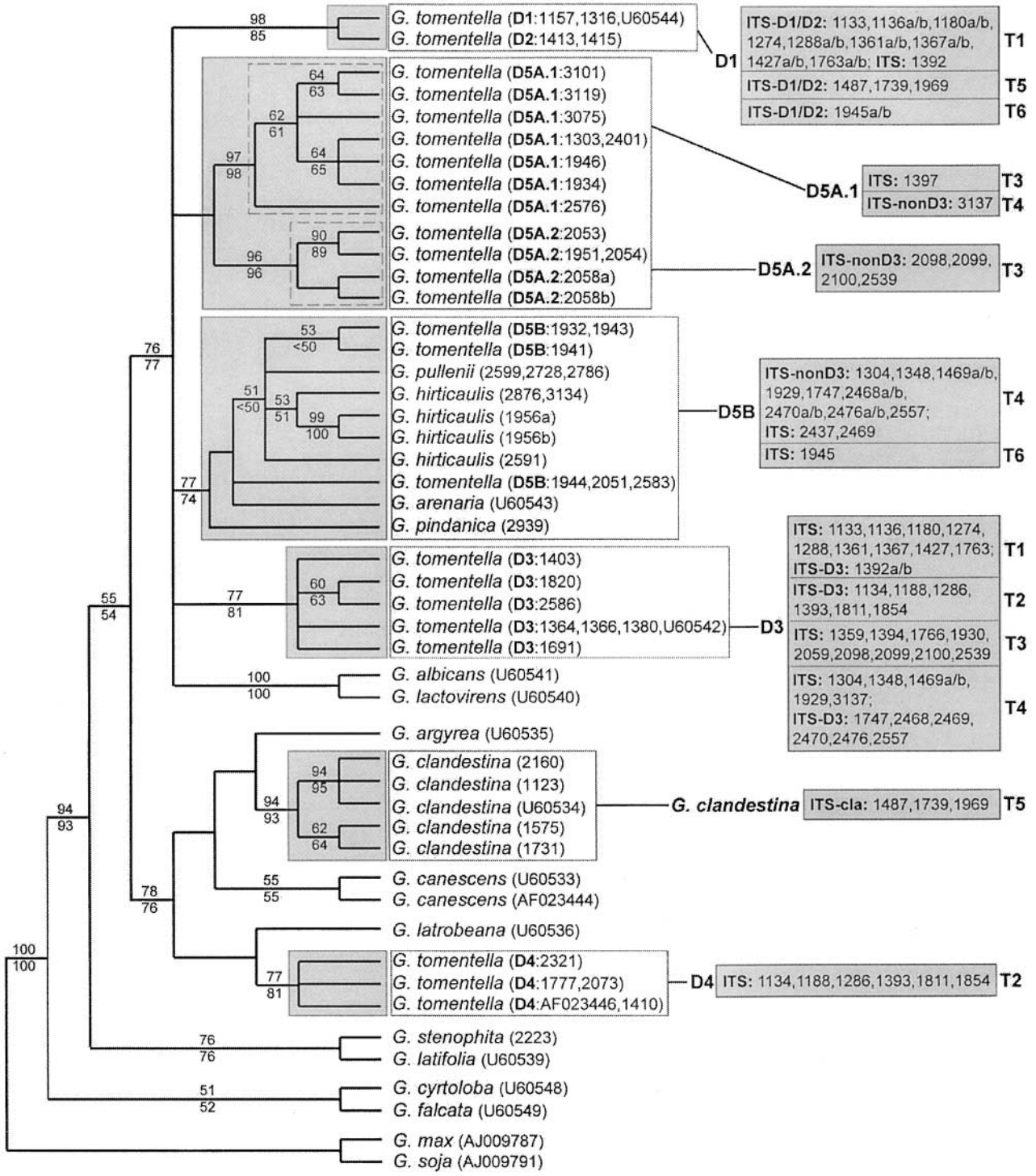

FIGURE 1.—The strict consensus tree resulting from a maximum-parsimony analysis of all diploid ITS sequences. Shaded boxes on the tree indicate clades that contain polyploid homeologous sequences when the analysis included both diploid and polyploid sequences; exact relationships between diploid and polyploid sequences for each of these clades are illustrated in Figures 2–6. Boxes on the right show the accessions of each tetraploid race for which ITS sequences were recovered and their phylogenetic affinities. Also indicated are the primers used to amplify and sequence these homeologues ("ITS" indicates that universal primers were used; "ITS-D3" and "ITS-cla" are repeat-specific primers and "ITS-nonD3" is an exclusion primer). Numbers above and below the branches are bootstrap values for the analysis with only diploid sequences and with all sequences, respectively; branches without numbers had bootstraps <50%. Four-digit numbers correspond to CSIRO "G" accessions and include new sequences or those from RAUSCHER *et al.* (2002). All other sequences are labeled with their GenBank accession numbers and are derived from either KOLLIPARA *et al.* (1997) or HSING *et al.* (2001).

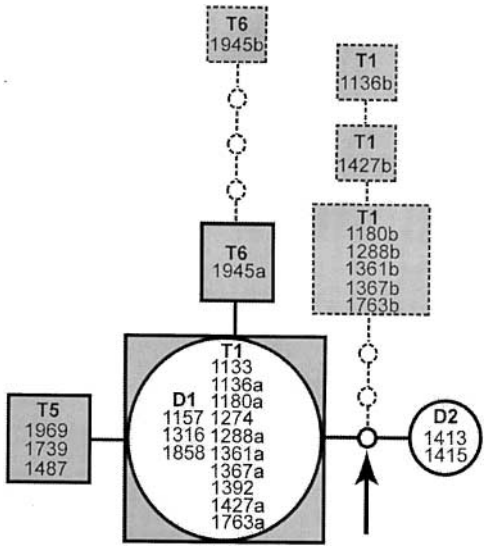

FIGURE 2.—ITS repeat network for all diploid D1 and D2 sequences and related polyploid homeologues. Shaded boxes and large open circles represent sequences from polyploid and diploid accessions, respectively; overlapping circles and boxes represent sequences that were identical in diploid and polyploid accessions. Numbers in the boxes or circles are CSIRO numbers for those accessions in which the sequence was detected (accession 1858 is U60544 from Figure 1). Lines indicate single mutational steps (either nucleotide substitutions or indels) and small circles are hypothetical intermediate repeat types that were not found. Dashed lines and boxes identify minor repeat sequences found in some accessions, inferred from polymorphic nucleotide sites on direct sequences from the ITS-D1/D2 specific primer (see text for details). Major and minor repeats from the same accession are labeled with "a" and "b," respectively. The arrow indicates the probable root of the network using other closely related clades as outgroups.

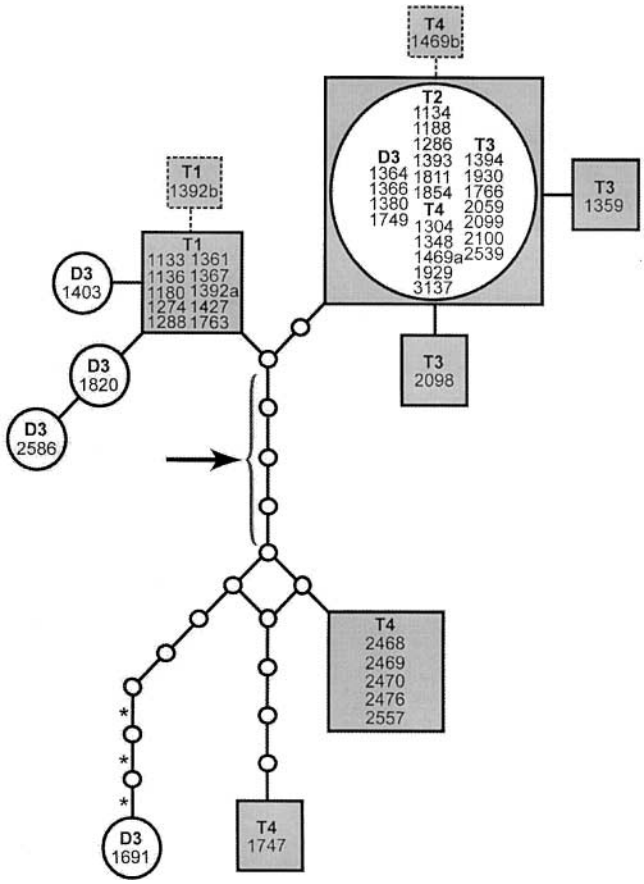

FIGURE 3.—ITS repeat network for all diploid D3 sequences and related polyploid homeologues. See Figure 2 for details on network symbols used. Asterisks next to lines show characters that were ambiguous in the polyploids due to missing data caused by the downstream position of the specific primer. Shown as autapomorphies of G1691, these characters may also be shared by the closely related T4 sequences and therefore belong on the long internal branch. The arrow indicates the probable root of the network, which is ambiguous depending on which taxa are used as outgroups. Accession 1749 is U60542 from Figure 1.

D1 repeat by a single nucleotide substitution (Figure 2); unlike the T1 or T6 accessions, these plants did not contain a detectable minor D1 repeat.

Four of the polyploid races (T1, T2, T3, and T4) contained D3 homeologues. There was no variation in the D3 repeat among any of the T1 accessions sampled, except for G1392, which had a minor repeat differing by 1 bp (Figure 3). Sequences from the T1 accessions were not identical to any sampled diploid accession, but were closely related to sequences from G1403 and G1820 (Figure 3). D3 homeologues from all of the T2 polyploids, from nearly all of the T3 polyploids, and from about half of the T4 polyploids were identical to the repeat of an extant sequence detected in four diploid accessions. The remaining two T3 accessions (G1359 and G2098) had D3 sequences that each differed from this common repeat type by a single autapomorphy. A second group of T4 polyploids (G1747, G2468, G2469, G2470, and G2476) had a D3 homeologue that was divergent from those of all other polyploids, including the remaining T4 accessions. Although they were not identical to any sampled diploid sequences, they were closely related to a D3 diploid accession that is notably divergent from all other diploids (G1691; Figure 3). The exact structure of the gene tree with respect to G1691 was ambiguous due to missing data in the polyploids that made it unclear whether three nucleotide characters were autapomorphies of G1691 or were shared with the polyploid sequences (Figure 3). There was also significant variation between G1747 and the other T4 accessions in this group.

Although there was some variation within the D4 clade, all sequences from the T2 polyploids were identical to the most common D4 diploid repeat (Figure 4). Similarly, analysis of the *G. clandestina* clade (which is closely related to the D4 clade; Figure 1) showed that all homeologous repeats from the T5 polyploids were either identical to or one mutational step away from one of the diploid accessions (G1126; Figure 4).

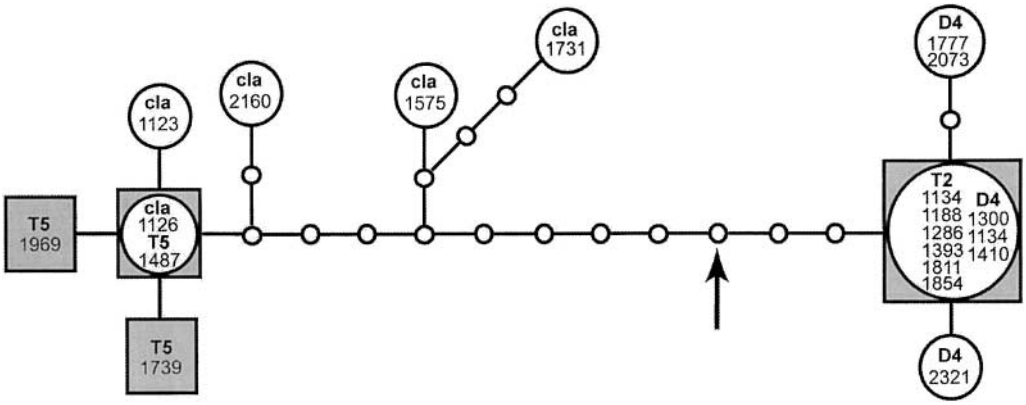

FIGURE 4.—ITS repeat network for all diploid D4 and *G. clandestina* sequences and related polyploid homeologues. See Figure 2 for details on network symbols used. Accession 1126 is U60534 and 1300 is AF02-3446 from Figure 1.

The D5A diploid sequences were highly variable and the phylogenetic analysis revealed two distinct clades, which have been labeled D5A.1 and D5A.2 (Figures 1 and 5). One T3 accession (G1397) had a D5A homeologue that was identical to an extant diploid sequence (G3119) in the D5A.1 clade. Also grouping with the D5A.1 accessions was the anomalous T4 (G3137), which was identical to the diploid accession G1934. The only other polyploid accessions for which the D5A repeat could be isolated were the collections from the island of Timor (G2098, G2099, G2100, and G2539) and these sequences grouped with the D5A.2 clade, but were not identical to any sampled diploid sequences. In the D5A gene tree, these sequences branch from a node connecting two branches of diploids (G2058 and G2054). Because the autapomorphic character forming the branch to G2058 (Figure 5) was among the missing data in the polyploid sequences, it is also possible that the polyploid accessions share this character.

Both the T4 and T6 tetraploids contained D5B homeologues (Figure 6). One T4 accession (G2437) was identical to a diploid sequence (G1941), while all other accessions were not identical to diploid sequences and grouped in two different clades on the tree. The D5B homeologue from the T6 accession was not identical to, but grouped one step away from sequences from a closely related species, *G. pullenii*, a member of the H genome that has previously been shown to have close affinities with the *G. tomentella* D5B group (BROWN *et al.* 2002).

DISCUSSION

nrDNA homeologue evolution in Glycine: The evolution of the nrDNA gene family is extremely complex and the mechanisms of concerted evolution are not well understood. Allopolyploids have the unique potential to help elucidate these mechanisms, as well as other genomic interactions such as nucleolar dominance, because their genomes are hybrid and, at least in the first generation, contain a minimum of two unique, homeologous ribosomal DNA loci. Describing the fate of these loci following the polyploid event is a necessary first step in understanding mechanisms of rDNA evolution, and the *G. tomentella* polyploid complex provides an interesting opportunity to study this because it includes at least six species that are derived from a common set of progenitor diploid species.

Direct sequencing of the ITS region using universal primers is a useful tool for obtaining a qualitative estimate of the relative ratio of homeologous nrDNA repeats from the two progenitor species that exist in an allopolyploid. A previous study (RAUSCHER *et al.* 2002), which used ITSs amplified from two diploids and mixed

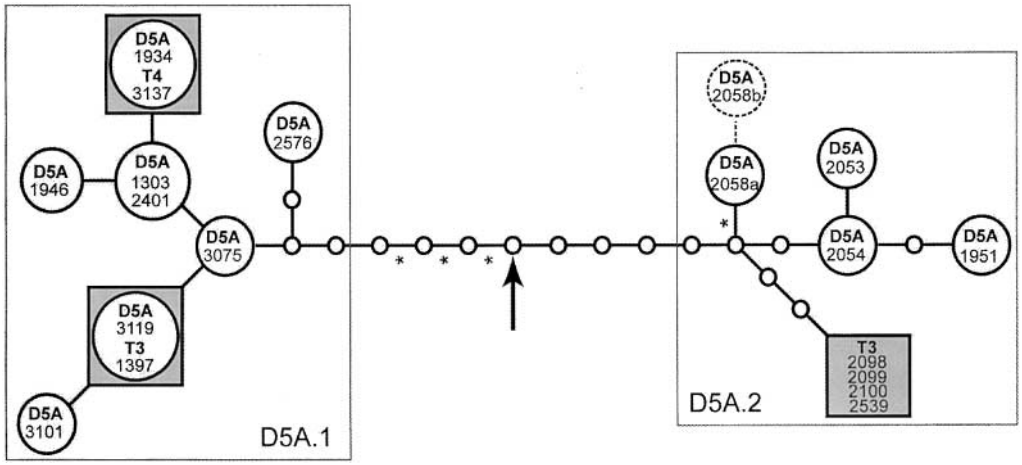

FIGURE 5.—ITS repeat network for all diploid D5A sequences and related polyploid homeologues. See Figure 2 for details on network symbols used. Large boxes circumscribe the two divergent D5A clades. Asterisks next to lines show characters that were ambiguous in the polyploids due to missing data caused by the downstream position of the ITS non-D3 exclusion primer.

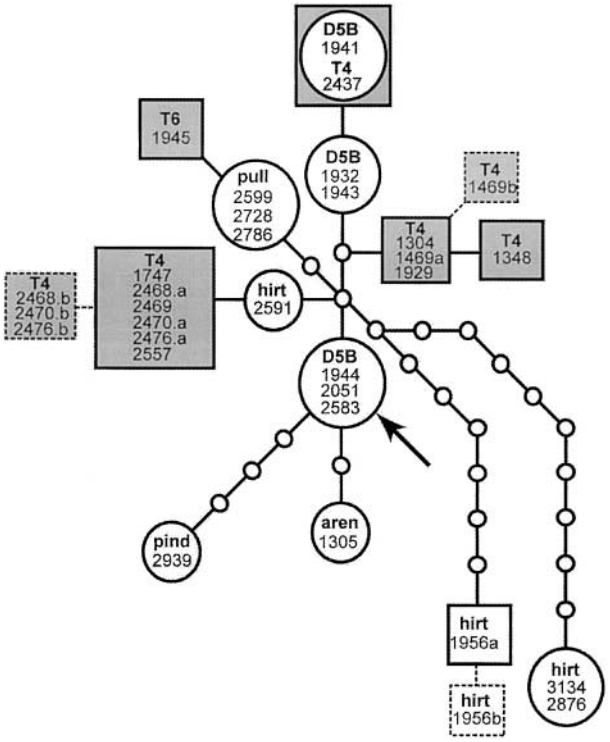

Figure 6.—ITS repeat network for all D5B and related diploid sequences and associated polyploid homeologues. See Figure 2 for details on network symbols used. The open box around one accession of *G. hirticaulis* indicates that it is a polyploid, but not part of the complex studied in this article. pull, *G. pullenii*; pind, *G. pindanica*; aren, *G. arenaria*; hirt, *G. hirticaulis*. Accession 1305 is U60543 from Figure 1.

at known concentrations, showed that direct sequencing (at least using this particular sequencing chemistry) will probably reveal signal only for sequences that are at least 10% or more of the population of PCR amplicons. Therefore, the fact that 34 of the 42 accessions included in this study produced clean ITS sequence (Table 1) suggests that in the genomes of most of these allopolyploids there is a bias of at least 9:1 in favor of one nrDNA homeologue.

Two possible alternative explanations for bias toward one homeologue are PCR selection and PCR drift (Wagner *et al.* 1994). The former, which could result from differences in primer affinities between homeologues, is unlikely since the same primers favored different repeats in different accessions, even within the same polyploid race (*e.g.*, T1, T3, and T4; Table 1). PCR drift, a consequence of random events early in the cycling, is also unlikely to account for most of this variation. A previous experiment showed that estimates of the ratio of the two homeologous repeats derived from direct sequencing were consistent with those obtained from screening clones of PCR products from independent reactions (Rauscher *et al.* 2002). In addition, homeologue repeat ratios determined by direct sequencing have been shown to be repeatable for different extractions from the same plant, for both rDNA and rtPCR of rRNA (S. Jolie, S. Sherman-Broyles, J. T. Rauscher, J. J. Doyle and A. H. D. Brown, unpublished results).

The imbalance in relative copy number was most extreme in the single T4 and several T3 accessions for which a second homeologous sequence could not be amplified with specific primers and in the T1 accessions that consistently had very weak D1 amplification (Rauscher *et al.* 2002). While some of this variation might be explained by differences in the number of nrDNA copies contributed by the two diploid ancestors that gave rise to these polyploids, it is more likely, especially in accessions with missing or weakly amplifying homeologues, that there has been a loss or homogenization of nrDNA repeats following the polyploid event.

These results in Glycine are similar to those obtained from direct sequencing of the ITS region in allopolyploids of the genus Gossypium (Wendel *et al.* 1995) in which monomorphic sequences were also observed. In the case of Gossypium, the authors suggested that interlocus concerted evolution was responsible for the observed homogenization of nrDNA tandem arrays in the polyploids. This inference was possible because *in situ* hybridization data were available that demonstrated the presence of four nrDNA loci in Gossypium allopolyploids (diploids have two loci), ruling out the complete loss of a homeologous locus. In addition, estimates of the number of nrDNA repeat units in the polyploids were additive relative to those of diploid progenitors, again precluding the loss of repeats as an explanation for the observed pattern (Wendel *et al.* 1995).

In Glycine, inferring interlocus concerted evolution as the mechanism responsible for bias in nrDNA homeologue copy number is more problematic since limited data are available on the structure and number of loci. Among the available data are two previous studies (Krishnan *et al.* 2001; Singh *et al.* 2001) that used fluorescence *in situ* hybridization (FISH) to estimate the number of NOR loci in several species in the genus, including a few accessions of *G. tomentella* polyploids. In at least one case, the results seem to be consistent with interlocus concerted evolution. Two NOR loci were detected (Singh *et al.* 2001) in a T3 polyploid accession (G1359; cited as IL446988) for which only a single ITS repeat (the D3 homeologue) could be isolated in this study, even with the use of specific primers (diploids have one locus). Since the specific primers are sensitive enough to detect rare repeats in a ratio of at least 1:1000 (Rauscher *et al.* 2002), it seems likely that the second locus has been partially or completely converted to the D3 homeologue. Without estimates of absolute copy number, however, this remains speculative.

In contrast to the T3 accession, a T1 polyploid that was studied using FISH demonstrated only a single NOR locus (Singh *et al.* 2001), suggesting to the authors that one homeologous locus had been silenced or lost. While

the specific accession used in the FISH study was not available for our ITS study, PCR with specific primers was successful in isolating both ITS homeologues from all T1 accessions examined. This can be explained either by the presence of a much reduced second locus that was not detected by FISH or by a single locus composed predominantly of one homeologous repeat but with <10% of the second homeologous repeat class. The former explanation seems more plausible.

Although it is unclear whether the conversion of nrDNA repeats through interlocus concerted evolution or the loss of nrDNA repeats is responsible for biases in *G. tomentella* homeologue ratios, two clear patterns are evident. First, the bias is not always toward the same repeat type. For example, the D3 homeologue is predominant in most accessions of the T1 (D1 × D3) and T3 (D3 × D5A) polyploids, but not in the T2 (D3 × D4) polyploids (Table 1). This result is similar to the condition in Gossypium, in which concerted evolution was found to favor the A genome in some allopolyploid species and the D genome in others (Wendel *et al.* 1995). In Glycine, however, this pattern is even more striking in that different homeologues have been favored among different accessions of the same polyploid race. For example, the D1 repeat was predominant in a single accession (G1392) of T1 polyploid while in all other T1 accessions sampled the D3 was predominant (Table 1). Similarly, in the T3 polyploids the D3 homeologue was predominant in all accessions but one (G1397; Table 1). The T4 polyploids were particularly diverse in this respect, with some accessions favoring one homeologue, others favoring the second, and still others with additive sequence. This diversity in patterns of nrDNA homeologue-copy-number bias may be common in other polyploid taxa as well, but requires thorough sampling across broad geographical areas to detect.

The second pattern of note was the fact that in most of these allopolyploids homogenization of the nrDNA repeats has not gone to completion. In contrast to previous studies of ITSs in *G. tomentella* (Hsing *et al.* 2001), both of the parental homeologous sequences were isolated from 36 of the 42 accessions sampled, even from accessions for which direct sequencing with universal primers showed no trace of a second repeat. In most Glycine polyploids the history of hybrid origin has been retained in the nrDNA, providing a useful source of phylogenetic information. Obtaining this information from rare repeats, however, requires a method that not only detects them, but also makes them available for sequencing. Southern hybridization has been used in groups such as Gossypium to detect rare nrDNA repeats in polyploids (Wendel *et al.* 1995), but requires diagnostic restriction site differences between homeologues and does not provide a practical way to isolate those repeats for sequencing. PCR with specific and exclusion primers not only is an effective tool for both detecting and sequencing rare homeologous ITS repeats in polyploids, but also is more practical for screening large numbers of accessions (Rauscher *et al.* 2002).

Origins of *G. tomentella* allopolyploids: Because allopolyploid species are the result of hybridization events, it is possible for the same species to originate multiple times, in different geographical areas and at different points in time. Having information on the number and timing of these events is useful, not only for understanding the evolution of a particular group of plants, but also for interpreting genomic variation between accessions of a polyploid species.

Phylogenetic analysis of ITS sequences from diploids and both homeologous copies of the polyploids can help to identify diploid progenitors as well as elucidate the time of origin and number of independent origins of these polyploids. In nearly all *G. tomentella* polyploid accessions for which two ribosomal repeats could be amplified and sequenced, phylogenetic analysis of the ITS homeologues confirmed the results of previous studies that identified the diploid progenitors (Doyle *et al.* 2002). Thus, in *G. tomentella*, ITSs seem to be a reliable tool for identifying diploid progenitors of allotetraploids. The only exception was a single T4 accession (G3137) that, in addition to its D3 ITS homeologue, had a D5A homeologue instead of the D5B homeologue, which had been found with nuclear histone H3D sequence data (Doyle *et al.* 2002); additional research will be necessary to explain this incongruence.

When rapidly evolving homeologous alleles in polyploids are identical or nearly identical to diploid alleles, the polyploid is assumed to have arisen relatively recently. For most polyploid accessions in *G. tomentella*, ITS repeat sequences were identical to those found in extant diploids. Several other polyploid sequences had single-base-pair autapomorphies distinguishing them from diploid repeats, including the D3 homeologue of two T3 accessions (G1359 and G2098; Figure 3), the D1 homeologue of all three T5 accessions (Figure 2), and both the D1 and D5B homeologues from the T6 (Figures 2 and 6). The low level of divergence between diploid and polyploid sequences is consistent with previous studies of the histone H3D nuclear gene in both the *G. tomentella* (Doyle *et al.* 2002) and the *G. tabacina* polyploid complexes (Doyle *et al.* 1999) and provided an estimate of the origin of the latter species of <30,000 years ago (Doyle *et al.* 1999, 2002).

Several polyploid accessions, however, had ITS sequences that were not identical to any sampled diploids. This was especially true in many of the T3 (D3 × D5A) and T4 (D3 × D5B) polyploids. The highest level of divergence was in the D3 homeologues from one group of T4 accessions (*e.g.*, G1747, G2568, and G2476; Figure 3), which differed by at least six to eight mutational steps from the most closely related diploid (G1691). At the histone H3D locus, a similar result was discovered, with four to seven mutations between the polyploid and diploid D3 sequences (Doyle *et al.* 2002). While the

possibility of a considerably more ancient polyploid event cannot be ruled out, it seems more likely that this divergence is the result of the extinction or lack of sampling of diploid D3 individuals that are closely related to the progenitors of the polyploid. The fact that only one diploid individual (G1691) of this divergent D3 clade has been discovered to date may attest to its current rarity.

Due to the fact that nrDNA loci are organized in tandem repeats, it is possible that, in addition to variation between homeologous loci, there is also variation between repeats within a locus, a pattern that has been found in other polyploid species (*e.g.*, Hughes *et al.* 2002) and that may provide additional clues on polyploid origins and evolution. In most Glycine polyploids, direct sequencing of a single parental ITS homeologue with either universal or specific primers revealed no observable repeat variation that could be detected as double peaks on the electropherogram, suggesting that minor variants or pseudogenes are in low copy number relative to the major repeat type. Two exceptions to this pattern were the D1 repeats from the T1 and T6 polyploids. In these accessions, the D1 sequence was amplified and isolated using the D1/D2-specific primer, and in both races signals from major and minor D1 repeats were detected. In the T1 polyploids, the second repeat does not seem to be a descendant of the dominant D1 repeat, but instead forms a clade that diverges from an internal node on the D1 network (Figure 2). The fact that this variation is found in nearly all of the T1 polyploids and across a wide geographic area suggests either that it was present in the diploid D1 ancestral population prior to the polyploid event, as has been suggested for other polyploid taxa (Hughes *et al.* 2002) or that it arose rapidly in a polyploid ancestor that gave rise to the modern T1 race. Evidence for the first hypothesis would require finding a D1 diploid that contained this minor repeat, but in the limited D1 accessions sampled to date it has not been found. The lack of a minor repeat in the two accessions from Brampton Island (G1133 and G1274) and the one accession in which the D1 repeat is predominant (G1392) may point to independent origins of these T1 polyploids or independent evolutionary histories that resulted in the loss or decreased relative copy number of the minor D1 repeat.

Phylogenetically informative variation among accessions of the same tetraploid race, in either homeologous parental copy, is useful for detecting independent origins of the race. Polyploids that group with different diploids on the gene tree for either of their two homeologues are most likely the result of independent origins. Autapomorphic variation between polyploid accessions could result either from independent origins from diploids that have not been sampled or from divergence following a single origin and so gives little support for multiple origins in the absence of additional evidence.

In many cases for which multiple accessions were available from the same *G. tomentella* polyploid race, there was little to no ITS sequence variation in either homeologous copy, precluding any test of multiple origins. As was found with histone H3D data (Doyle *et al.* 2002), there was no variation among T2 or T5 accessions other than single-base-pair autapomorphies, and the only variation among T1 accessions was between the major and minor D1 repeats. As mentioned above, it is possible that the presence of a minor D1 repeat in some T1 accessions and not in others is a result of different origins. Similarly, the fact that accession G1392 has an unusual D1/D3 homeologue copy ratio, demonstrated by the amplification of D1, not D3 repeats with universal primers, may suggest another independent origin of the T1 race. However, this may also have been caused by differential patterns of concerted evolution. In both of these cases, additional data from other loci are necessary to discriminate between these hypotheses.

In the T3 polyploids, the D3 homeologue was nearly identical in all accessions, except for two with single-base-pair autapomorphies. Unfortunately, since the D5A homeologue was absent in several of the T3 accessions, it is not possible to assess diversity across the entire polyploid race, but at least one accession, G1397, had a D5A repeat that differed significantly from the D5A repeat in all other T3 accessions. In fact, the diploid gene tree shows that there are two distinct D5A clades, and the T3 repeats fall out in each of these two clades. This result is nearly identical to that previously found using the histone H3D locus (Figure 5 in Doyle *et al.* 2002), in which two distinct D5A clades were also identified; G1397 grouped with the D5A.1 diploid accessions and all others grouped with the D5A.2 diploids. The four T3 accessions for which no D5A ITS repeat could be found in this study (G1359, 1394, 1766, and 2059) belonged to a divergent branch of the D5A.2 clade on the histone H3D gene tree. But because this branch coalesced to the same diploid sequence on the histone tree as the other T3 accessions (those from the island of Timor, including G2098, 2099, 2100, and 2539), no clear inference of multiple origins could be made (Doyle *et al.* 2002). The fact that these two groups have very different patterns of ITS homeologue evolution, however, may support the hypothesis of their separate origin. If so, this would suggest that there have been at least three independent origins of the T3 polyploids, represented by G1397, the Timor accessions, and the group lacking a D5A nrDNA homeologue.

The T4 polyploids showed an even more striking level of between-accession diversity for both of the ITS homeologues. In the D5B gene tree, T4 repeats are found in three distinct clades (Figure 6). Four accessions from Northern Territory, western Queensland, the Philippines, and Taiwan (G1304, 1348, 1469, and 1929) group in a unique clade, six more from the Cape York Peninsula of northern Queensland (G1747, 2468, 2469, 2470,

2476, and 2557) group with an accession of *G. hirticaulis* (a close relative of the *G. tomentella* D5B accessions), and a single accession from Western Australia (G2437) is identical to a diploid D5B accession. This level of diversity is found in the D3 homeologues as well (Figure 3). The first two groups of T4 accessions identified on the D5B tree also group on two divergent clades of the D3 tree, one of which is identical to several sampled diploid accessions (*e.g.*, G1364) and the other of which groups with a unique diploid accession, G1691. Among the Cape York Peninsula polyploids, G1747 was divergent from the others, but this variation was autapomorphic and so provided no strong evidence for an additional origin. The last accession, G2437, had no detectable D3 repeat. Together, the two ITS gene trees (D3 and D5B) strongly suggest at least three independent origins of the T4 polyploids. These results are in agreement with the histone allele networks, especially for the D5B alleles, which found the same three distinct groups (Doyle *et al.* 2002, Figures 5 and 7).

These results show that sequence data from nrDNA ITS homeologous loci have the potential to provide significant insights into the origins of allopolyploid species. In Glycine, evidence from multiple loci, including two ITS homeologues and two histone H3D homeologues, is beginning to elucidate a complex pattern of evolution, which shows not only that the *G. tomentella* complex is composed of several different species that combine different parental genomes, but also that several of these races have arisen multiple times. In the future it will be instructive to compare these gene trees to those of other loci, especially from the chloroplast, to see if there is congruence or evidence for even more origins of these allopolyploid races. Multiple origins have been found in several polyploid plant taxa such as Tragopogon (Soltis and Soltis 1991; Cook *et al.* 1998), Heuchera (Segraves *et al.* 1999), and Asplenium (Trewick *et al.* 2002), and even in some animals (Tsigenopoulos *et al.* 2002), suggesting that a complete understanding of evolution in any allopolyploid taxon requires broad geographical sampling of multiple accessions and phylogenetic data from multiple loci. Together, the variation in patterns of nrDNA molecular evolution and multiple origins of polyploids elucidated in this and previous studies will, in the future, serve as a useful base with which to study genome interactions (such as nucleolar dominance) and patterns of gene expression in this diverse allopolyploid complex.

The authors acknowledge Jane Doyle for her assistance with work in the laboratory and Marc Ellis, Rohit Mago, and Randy Bayer for valuable comments on a previous version of this manuscript. This research was funded by grant DEB-0098483 from the U.S. National Science Foundation Systematic Biology Program.

LITERATURE CITED

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell *et al.*, 1995 The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. **82:** 247–277.

Bretagnolle, F., F. Felber, F. G. Calame and P. Kupfer, 1998 Polyploidy in plants. Bot. Helv. **108:** 5–37.

Brown, A. H. D., J. L. Doyle, J. P. Grace and J. J. Doyle, 2002 Molecular phylogenetic relationships within and among diploid races of Glycine tomentella (Leguminosae). Aust. Syst. Bot. **15:** 37–47.

Buckler, E. S., A. Ippolito and T. P. Holtsford, 1997 The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics **145:** 821–832.

Chen, Z. J., L. Comai and C. S. Pikaard, 1998 Gene dosage and stochastic effects determine the severity and direction of uniparental rRNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. Proc. Natl. Acad. Sci. USA **95:** 14891–14896.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1659.

Cook, L. M., P. S. Soltis, S. G. Brunsfeld and D. E. Soltis, 1998 Multiple independent formations of *Tragopogon* tetraploids (Asteraceae): evidence from RAPD markers. Mol. Ecol. **7:** 1293–1302.

Crandall, K. A., 1994 Intraspecific cladogram estimation—accuracy at higher levels of divergence. Syst. Biol. **43:** 222–235.

Doyle, J. J., and A. H. D. Brown, 1989 5S nuclear ribosomal gene variation in the *Glycine tomentella* polyploid complex (Leguminosae). Syst. Bot. **14:** 398–407.

Doyle, J. J., J. L. Doyle and A. H. D. Brown, 1999 Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. Proc. Natl. Acad. Sci. USA **96:** 10741–10745.

Doyle, J. J., J. L. Doyle, A. H. D. Brown and R. G. Palmer, 2002 Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. Evolution **56:** 1388–1402.

Doyle, M. J., and A. H. D. Brown, 1985 Numerical analysis of isozyme variation in *Glycine tomentella*. Biochem. Syst. Ecol. **13:** 413–419.

Doyle, M. J., J. E. Grant and A. H. D. Brown, 1986 Reproductive isolation between isozyme groups of *Glycine tomentella* (Leguminosae), and spontaneous doubling in their hybrids. Aust. J. Bot. **34:** 523–535.

Hamby, R. K., and E. A. Zimmer, 1992 Ribosomal RNA as a phylogenetic tool in plant systematics, pp. 50–91 in *Molecular Systematics of Plants*, edited by P. S. Soltis, D. E. Soltis and J. J. Doyle. Chapman & Hall, New York.

Hsing, Y. C., J.-S. Hsieh, C.-I. Peng, C.-H. Chou and T.-Y. Chiang, 2001 Systematic status of the *Glycine tomentella* and *G. tabacina* species complexes (Fabaceae) based on ITS sequences of nuclear ribosomal DNA. J. Plant Res. **114:** 435–442.

Hughes, C. E., C. D. Bailey and S. A. Harris, 2002 Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. Am. J. Bot. **89:** 1057–1073.

Kollipara, K. P., R. J. Singh and T. Hymowitz, 1994 Genomic diversity and multiple origins of tetraploid (2n=78, 80) *Glycine tomentella*. Genome **37:** 448–459.

Kollipara, K. P., R. J. Singh and T. Hymowitz, 1997 Phylogenetic and genomic relationships in the genus *Glycine* Willd. based on sequences from the ITS region of nuclear nrDNA. Genome **40:** 57–68.

Krishnan, P., V. T. Sapra, K. M. Soliman and A. Zipf, 2001 FISH mapping of the 5S and 18S–28S nrDNA loci in different species of Glycine. J. Hered. **92:** 295–300.

Lee, H. S., and Z. J. Chen, 2001 Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc. Natl. Acad. Sci. USA **98:** 6753–6758.

Leitch, I. J., and M. D. Bennett, 1997 Polyploidy in angiosperms. Trends Pharmacol. Sci. **2:** 470–476.

Lewis, M. S., and C. S. Pikaard, 2001 Restricted chromosomal silencing in nucleolar dominance. Proc. Natl. Acad. Sci. USA **98:** 14536–14540.

Maddison, W. P., and D. R. Maddison, 1992 *MacClade*. Sinauer Associates, Sunderland, MA.

Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. Trends Genet. **19:** 141–147.

Pikaard, C. S., 2000 The epigenetics of nucleolar dominance. Trends Genet. **16:** 675–677.

Pikaard, C. S., 2001 Genomic change and gene silencing in polyploids. Trends Genet. **17:** 675–677.

Rambaut, A., 1995 *Se-Al. Sequence Alignment Program.* University of Oxford, Oxford.

Rauscher, J. T., J. J. Doyle and A. H. D. Brown, 2002 Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol. Ecol. **11:** 2691–2702.

Sang, T., D. J. Crawford and T. F. Stuessy, 1995 Documentation of reticulate evolution in peonies (Peonia) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. Proc. Natl. Acad. Sci. USA **92:** 6813–6817.

Segraves, K. A., J. N. Thompson, P. S. Soltis and D. E. Soltis, 1999 Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia.* Mol. Ecol. **8:** 253–262.

Singh, R. J., K. P. Kollipara and T. Hymowitz, 1987 Polyploid complexes of *Glycine tabacina* (Labill.) Benth. and *G. tomentella* Hayata revealed by cytogenetic analysis. Genome **29:** 490–497.

Singh, R. J., H. H. Kim and T. Hymowitz, 2001 Distribution of nrDNA loci in the genus *Glycine* Willd. Theor. Appl. Genet. **103:** 212–218.

Soltis, D. E., and P. S. Soltis, 1991 Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): nrDNA evidence. Syst. Bot. **16:** 407–413.

Soltis, D. E., and P. S. Soltis, 1999 Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. **14:** 348–352.

Song, K., P. Lu, K. Tang and T. C. Osborn, 1995 Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92:** 7719–7723.

Swofford, D. L., 2000 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sinauer Associates, Sunderland, MA.

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

Trewick, S. A., M. Morgan-Richards, S. J. Russell, S. Henderson, F. J. Rumsey *et al.*, 2002 Polyploidy, phylogeography and Pleistocene refugia of the rockfern *Asplenium ceterach*: evidence from chloroplast DNA. Mol. Ecol. **11:** 2003–2012.

Tsigenopoulos, C. S., P. Ráb, D. Naran and P. Berrebi, 2002 Multiple origins of polyploidy in the phylogeny of southern African barbs (Cyprinidae) as inferred from mtDNA markers. Heredity **88:** 466–473.

Vazquez, M. L., 2001 Molecular and morphological studies on Mexican red oaks (*Quercus* sect. *Lobatae*). Ph.D. Thesis, Cornell University, Ithaca, NY.

Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof *et al.*, 1994 Surveys of gene families using polymerase chain-reaction—PCR selection and PCR drift. Syst. Biol. **43:** 250–261.

Wendel, J. F., 2000 Genome evolution in polyploids. Plant Mol. Biol. **42:** 225–249.

Wendel, J. F., A. Schnabel and T. Seelanan, 1995 Bi-directional interlocus concerted evolution following allopolyploid speciation in cotton (Gossypium). Proc. Natl. Acad. Sci. USA **92:** 280–284.

White, T. J., T. Bruns, S. Lee and J. Taylor, 1990 Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenies, pp. 315–322 in *PCR Protocols*, edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White. Academic Press, San Diego.

Communicating editor: S. W. Schaeffer