Contrasting Modes of Diversification in the *Aux/IAA* and *ARF* Gene Families[1][w]

David L. Remington*, Todd J. Vision, Thomas J. Guilfoyle, and Jason W. Reed

Department of Biology, University of North Carolina, Greensboro, North Carolina 27402–6170 (D.L.R.); Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599–3280 (T.J.V., J.W.R.); and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211 (T.J.G.)

The complete genomic sequence for Arabidopsis provides the opportunity to combine phylogenetic and genomic approaches to study the evolution of gene families in plants. The *Aux/IAA* and *ARF* gene families, consisting of 29 and 23 loci in Arabidopsis, respectively, encode proteins that interact to mediate auxin responses and regulate various aspects of plant morphological development. We developed scenarios for the genomic proliferation of the *Aux/IAA* and *ARF* families by combining phylogenetic analysis with information on the relationship between each locus and the previously identified duplicated genomic segments in Arabidopsis. This analysis shows that both gene families date back at least to the origin of land plants and that the major *Aux/IAA* and *ARF* lineages originated before the monocot-eudicot divergence. We found that the extant *Aux/IAA* loci arose primarily through segmental duplication events, in sharp contrast to the *ARF* family and to the general pattern of gene family proliferation in Arabidopsis. Possible explanations for the unusual mode of *Aux/IAA* duplication include evolutionary constraints imposed by complex interactions among proteins and pathways, or the presence of long-distance cis-regulatory sequences. The antiquity of the two gene families and the unusual mode of *Aux/IAA* diversification have a number of potential implications for understanding both the functional and evolutionary roles of these genes.

The complete Arabidopsis genomic sequence (The Arabidopsis Genome Initiative, 2000) has opened new avenues for understanding the composition, structure, organization, and evolution of a plant genome. One opportunity it affords is the ability to identify the sequence and genomic context of every member of a given gene family.

Despite the small size of the Arabidopsis genome (approximately 125 Mb), the majority of Arabidopsis genes belong to families containing two or more members. Some of this redundancy can be attributed to ancient, large-scale genomic duplications (Blanc et al., 2000, 2003; The Arabidopsis Genome Initiative, 2000; Vision et al., 2000). Well over half of the Arabidopsis genome is found in large duplicated blocks, which led to the early suggestion that Arabidopsis was an ancient tetraploid (Blanc et al., 2000; The Arabidopsis Genome Initiative, 2000). Some chromosomal regions, however, have multiple duplicates, and different pairs of regions appear to be of different ages, both of which argue for the occurrence of multiple independent duplication events (Ku et al., 2000; Vision et al., 2000; Simillion et al., 2002; Blanc et al., 2003). The most recent large-scale duplication event (almost certainly a genome-wide polyploidization event) occurred more recently than the Brassicaceae-Malvaceae divergence (Blanc et al., 2003; Bowers et al., 2003) approximately 81 to 94 million years ago (Mya; Wikstrom et al., 2001). Recent reports suggest that the duplication event may have occurred as recently as 40 Mya but is evidently older than the Arabidopsis-Brassica divergence (Simillion et al., 2002; Blanc et al., 2003; Bowers et al., 2003). Remnants of multiple older, large-scale duplication events have been identified (Vision et al., 2000; Simillion et al., 2002; Blanc et al., 2003; Bowers et al., 2003), the oldest of which predate the divergence of Arabidopsis and rice (*Oryza sativa*; Raes et al., 2003). Most duplicated genes currently found in Arabidopsis, however, appear to have resulted from numerous independent, small-scale duplication events (Vision et al., 2000), some of which produced tandem arrays of related genes while others produced dispersed gene families. A large number of such small-scale duplications have likely occurred since the most recent large-scale duplication, and it is reasonable to assume that most such duplications have since reverted to single copy (Lynch and Conery, 2000).

Two related gene families of interest in Arabidopsis are those coding for the Auxin Response Factor (ARF) and Aux/IAA proteins. The plant hormone auxin regulates development in all major land plant lineages and even the brown alga Fucus (Basu et al., 2002;

[1] This work was supported in part by the National Institutes of Health (Individual Postdoctoral Fellowship 5–F32–GM29554 to D.L.R. and grant no. R01–GM52456 to J.W.R.) and by the National Science Foundation (grant no. IBN–0116106 to J.W.R. and T.J.G., grant no. MCB–0080096 to T.J.G., and grant no. DBI–0227314 to T.J.V.).

* Corresponding author; e-mail dlreming@uncg.edu; fax 336–334–5839.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.104.039669.

Cooke et al., 2002). In angiosperms, auxin exerts its effect in part by inducing or repressing expression of numerous genes. ARF and Aux/IAA proteins are known to mediate auxin gene expression responses. Most ARF proteins have a conserved DNA-binding domain that recognizes auxin response elements (AuxREs) present in promoters; a middle domain that is highly divergent but, in all cases tested, has transcription activation or repression activity; and a C-terminal domain containing two motifs, called III and IV, that can mediate dimerization (Ulmasov et al., 1999a, 1999b; Hagen and Guilfoyle, 2002). The *Aux/IAA* gene family has been intensively studied in Arabidopsis and also to varying degrees in a number of other plants, including pea (*Pisum sativum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), cucumber (*Cucumis sativus*), and rice (Abel et al., 1995; Reed, 2001). Mutations in various family members have a variety of phenotypic effects on plant morphology and development (Reed, 2001). Completion of the Arabidopsis genome sequence has expanded the known complement of the *Aux/IAA* family to 29 loci and the *ARF* family to 23 loci.

Biochemical and genetic studies in Arabidopsis and other species have led to a working model for how these proteins mediate auxin responses (Gray et al., 2001; Tiwari et al., 2001, 2003; Hagen and Guilfoyle, 2002; Tian et al., 2003). In this model, ARF proteins bind to AuxREs in gene promoters and can either activate or repress transcription, depending on the middle domain they contain. When auxin levels are low, Aux/IAA proteins dimerize with ARF activators and thereby repress their activity. Auxin stimulates turnover of Aux/IAA proteins by increasing their interaction with the SCF^{TIR1} ubiquitin ligase, leading to their ubiquitination and degradation. This releases the ARFs from inhibition, allowing activation of gene expression. Auxin induces many genes encoding Aux/IAA proteins, and this model thus incorporates negative feedback loops. The model is based on study of just a few ARF and Aux/IAA proteins but provides a framework for understanding how multiple members of these families may function.

In this article, we combine a molecular phylogenetic analysis of the *Aux/IAA* and *ARF* families with information on genomic duplications in Arabidopsis in order to place the origin and proliferation of these two families with respect to the timing of major divergence and genomic duplication events in the Arabidopsis lineage. By comparing the complete set of *Aux/IAA* loci from Arabidopsis with sequences from other plants, we show that several extant lineages of *Aux/IAA* and *ARF* loci diverged long before the monocot-eudicot divergence and that the *Aux/IAA* family dates back at least to the origin of land plants. We show that surviving genes in the *Aux/IAA* family, but not the *ARF* family, arose predominantly through large-scale genomic duplication events. This unusual mode of diversification in the *Aux/IAA* family suggests several hypotheses, including the presence of unique functional constraints between members of this gene family and other unidentified loci or the presence of long-distance cis-regulatory sequences.

RESULTS

Phylogeny of the Arabidopsis *Aux/IAA* Family

We used neighbor-joining analysis to reconstruct phylogenetic trees of 28 Arabidopsis *Aux/IAA* translated sequences, excluding *IAA33*, which contains only portions of motifs III and IV. In order to identify the root position for the tree, motifs III and IV of seven representative *ARF* sequences were included as an outgroup under the assumption that the *Aux/IAA* and *ARF* families are sister to each other. The *Aux/IAA* and *ARF* sequences formed two separate clusters with moderate (63%) bootstrap support within the combined *IAA-ARF* neighbor-joining tree (data not shown). The deeper branches within the *Aux/IAA* cluster of the combined tree had poor bootstrap support, making the correct placement of the root uncertain. However, the root position between the *IAA32-34* cluster and the remaining *Aux/IAA* loci was identical in both neighbor-joining and strict consensus maximum parsimony trees. Moreover, *IAA32* and *IAA34* are among the most divergent *Aux/IAA* loci in their overall organization as well, as both loci lack motif II and a putative bipartite nuclear localization signal, and *IAA32* also lacks a recognizable motif I.

We analyzed the *Aux/IAA* family phylogeny in more detail, using all sites that could be aligned in at least some subsets of the family, including some sites outside the four conserved motifs. Only the *Aux/IAA* loci were included in these analyses, and *IAA32* and *IAA34* were treated as an outgroup to the remaining loci based on the root position inferred from the *IAA-ARF* alignment. Three analyses using alternate alignments of the more variable regions of the Aux/IAA protein produced only minor differences in the tree topology as described below, indicating that phylogenetic reconstruction was relatively insensitive to alignment uncertainties. The tree constructed from alignment 1 is shown in Figure 1. Fifteen of the *Aux/IAA* sequences were joined in a moderately well-supported branch, leading to node *e* in Figure 1, consistent with analyses by Abel et al. (1995) and Rogg et al. (2001). We subsequently refer to this set of loci as group A. The sister *IAA28* clade described by Rogg et al. (2001), however, is paraphyletic in our analysis, consisting of a nested set of subgroups basal to group A. We designate these sequences as group B for the sake of simplicity, while recognizing their apparent paraphyly. Three recently identified loci not included in the Rogg et al. (2001) analysis (*IAA29*, *IAA32*, and *IAA34*), were in turn basal to all other *Aux/IAA* loci, while three others (*IAA20*, *IAA30*, and *IAA31*) formed one of the nested subgroups within group B. Each of the nested subgroups of group B sequences consisted of three or four sequences with varying degrees of bootstrap

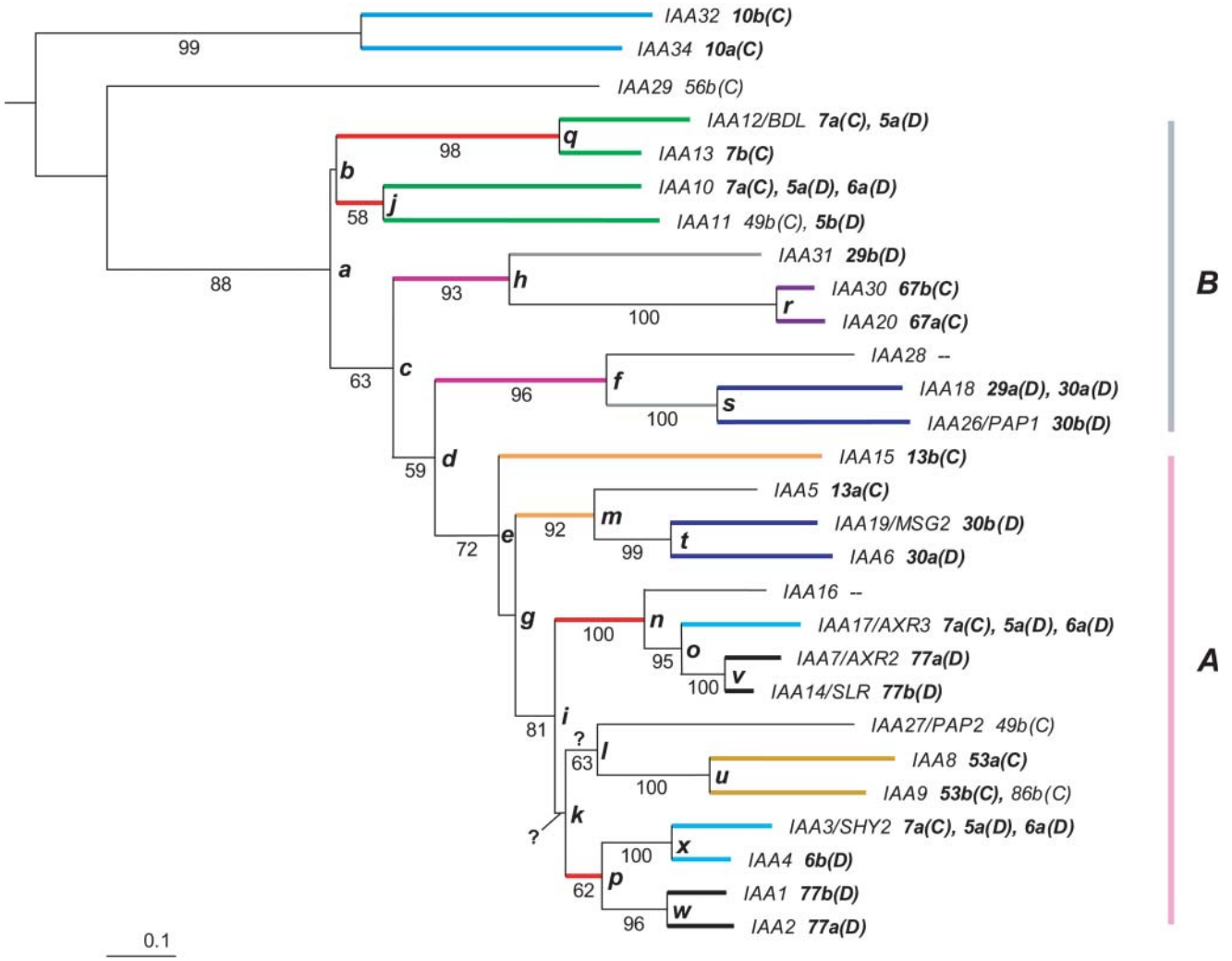

Figure 1. Neighbor-joining tree of Arabidopsis *Aux/IAA* sequences, using alignment 1 of three alternate alignments. The position of the root was determined from an outgroup consisting of seven *ARF* loci. The percent bootstrap support for 500 replicates is shown below each branch with >50% support. A question mark (?) denotes a branch that was not supported in trees constructed from alignment 2 and/or 3. All duplicated blocks per Vision et al. (2000), in which each sequence occurs, and the estimated age class (in parentheses) are listed after each locus name and are in bold if another *Aux/IAA* sequence is present in the block. Branches from the same putative chromosomal or local duplication are shown in the same color.

support for each. The nested topology of these subgroups had only weak to moderate support. Group A contained four subgroups of loci with varying degrees of bootstrap support, which are represented by nodes *l*, *m*, *n*, and *p* in Figure 1. Most of the relationships of the group A subgroups to one another and to *IAA15* were poorly resolved. The node *l* (*IAA8-9-27*) subgroup was placed as sister to the node *p* (*IAA1-2-3-4*) subgroup when alignment 1 was used, was basal to the node *n* (*IAA7-14-16-17*) subgroup with alignment 2, and was sister to the node *n* subgroup with alignment 3. Maximum parsimony methods were also used with alignments 1 and 3, and resulted in identical single minimum-length trees. The maximum parsimony trees were identical in topology to the neighbor-joining tree from alignment 1, except that *IAA10* was basal to *IAA11-12-13* in the maximum parsimony trees rather than sister to *IAA11* as in the neighbor-joining trees (data not shown).

Twenty of the 28 *Aux/IAA* loci formed 10 sister pairs in the neighbor-joining reconstructions, 9 of which had strong bootstrap support (≥96% in all three trees). Five pairs of sister loci (*IAA1* and *IAA2*; *IAA3/SHY2* and *IAA4*; *IAA6* and *IAA19/MSG2*; *IAA12/BDL* and *IAA13*; and *IAA20* and *IAA30*) are highly similar in stretches of their upstream flanking regions (Fig. 2). In the first four of these pairs, the regions of apparent homology contain multiple putative AuxREs, with matches of five out of six nucleotides or better to the consensus TGTCTC sequence (Ulmasov et al., 1999a, 1999b) in either forward or reverse orientation. Each of these conserved regions is located approximately 200 to

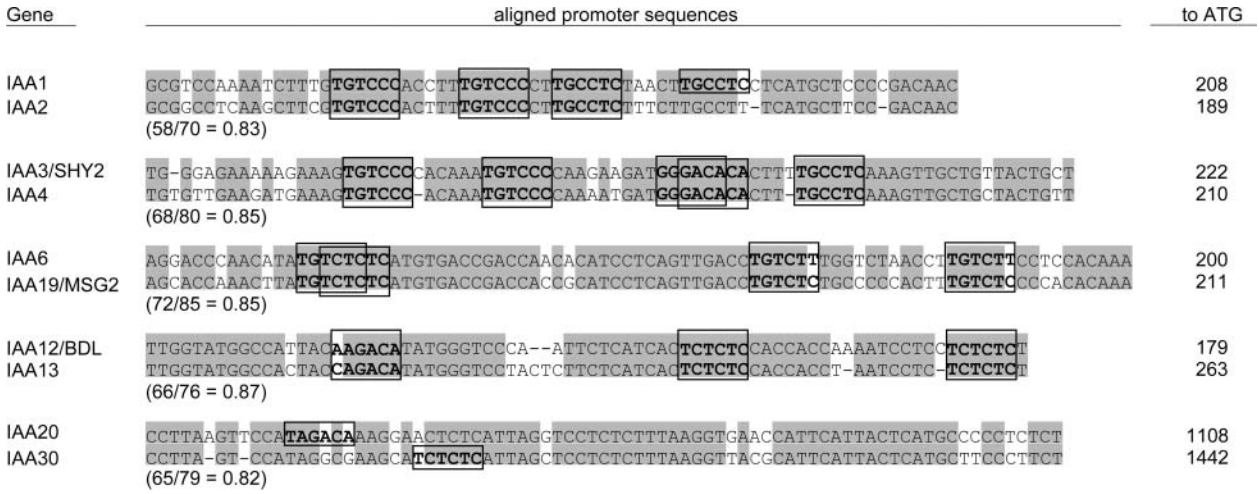

Figure 2. Alignments of conserved promoter regions for five sister pairs of Arabidopsis *Aux/IAA* loci. Identical bases are shaded, and the fraction and percentage of identity within the regions is given below each pair. The distance (bp) from the 3′ end of each region to the ATG translation start is shown at right. Motifs with at least 5/6 similarity in one or both sequences to the consensus AuxRE (TGTCTC/GAGACA) are shown in bold and boxed.

300 bp upstream of the start codon. By contrast, the region conserved between *IAA20* and *IAA30* is located much farther upstream and each of these loci contains only one potential and possibly spurious AuxRE in the conserved region.

Relationship of *Aux/IAA* Phylogeny to Chromosomal Duplications

All 10 of the sister locus pairs in Arabidopsis are located on homologous duplicated chromosomal segments identified by Vision et al. (2000; Figs. 1 and 3). Two nonsister sequence pairs (*IAA5* and *IAA15*, and *IAA18* and *IAA31*) are also located on homologous segments. Only four of the 28 *Aux/IAA* loci included in our analysis (*IAA27/PAP2*, *IAA16*, *IAA28*, and *IAA29*) lack any counterpart in a homologous segment, and three of these orphan loci show intriguing relationships to identified blocks. *IAA16* and *IAA28* are not located within any of the identified blocks but are positioned just beyond the corresponding ends of segments 72a and 72b, respectively. *IAA27/PAP2* is located on segment 49b, just beyond the identified terminus of segment 5b, which includes *IAA11*. Thus, it is possible that *IAA27/PAP2* is a descendant of the block 5 duplication, from which *IAA11* also arose.

Some blocks contain multiple sets of *IAA* genes, suggesting the occurrence of tandem or local duplications prior to the chromosomal block duplications. Block 30 contains pairs of both group A (*IAA6* and *IAA19/MSG2*) and group B (*IAA18* and *IAA26/PAP1*) genes. *IAA18* also falls within segment 29a, and *IAA31* (in a separate subgroup of group B) is in segment 29b. Blocks 29 and 30 appear to represent the duplication of a single ancestral chromosomal segment. Segments 29a and 30a overlap slightly, and segment 29b is immediately adjacent to segment 30b but is inverted relative to the 29a-30a orientation. Segment 29b may have been inverted and partially duplicated after the segmental duplication. Both blocks belong to the same inferred age classes (Vision et al., 2000; Blanc et al., 2003). The region of chromosome 1, encompassed by segments 5a, 6a, and 7a, has four sequences (*IAA3/SHY2* and *IAA17/AXR3* in group A, and *IAA10* and *IAA12/BDL* in group B). *IAA10*, *IAA12/BDL*, and *IAA3/SHY2* have sister duplicates in segments 5b (*IAA11*), 6b (*IAA13*), and 7b (*IAA4*), respectively, which lie on chromosomes 4, 5, and 2. Moreover, *IAA17/AXR3* and the *IAA3/SHY2-IAA4* sister pair are sibling to the *IAA7/AXR2-IAA14/SLR* and *IAA2-IAA1* sister pairs, respectively, each located in block 77. *IAA3/SHY2* and *IAA17/AXR3* are immediately adjacent to each other, as are *IAA1* and *IAA14/SLR*, and *IAA2* and *IAA7/AXR2* are separated by only one predicted open reading frame that lacks experimental confirmation as an expressed gene (Fig. 3). This pattern provides evidence of multiple rounds of tandem duplication prior to the segmental duplications giving rise to blocks 5, 6, and 7 and that giving rise to block 77.

The occurrence and patterns of duplicated blocks among the *Aux/IAA* loci provided the opportunity to map possible chromosomal duplication scenarios onto the gene family phylogeny. One such scenario (hereafter referred to as the base reconstruction), which excludes the three basal loci *IAA29*, *IAA32*, and *IAA34*, is presented in Figure 4. This scenario assumes that the neighbor-joining topology in Figure 1 accurately reflects the order of gene duplication events, and also assumes (1) a block 72 origin for *IAA16* and *IAA28*, (2) that segment 5b can be extended to contain *IAA27/PAP2*, and (3) that blocks 29 and 30 represent the same duplication event. The base reconstruction requires 20 separate tandem, block, and/or individual duplication events, including five separate tandem

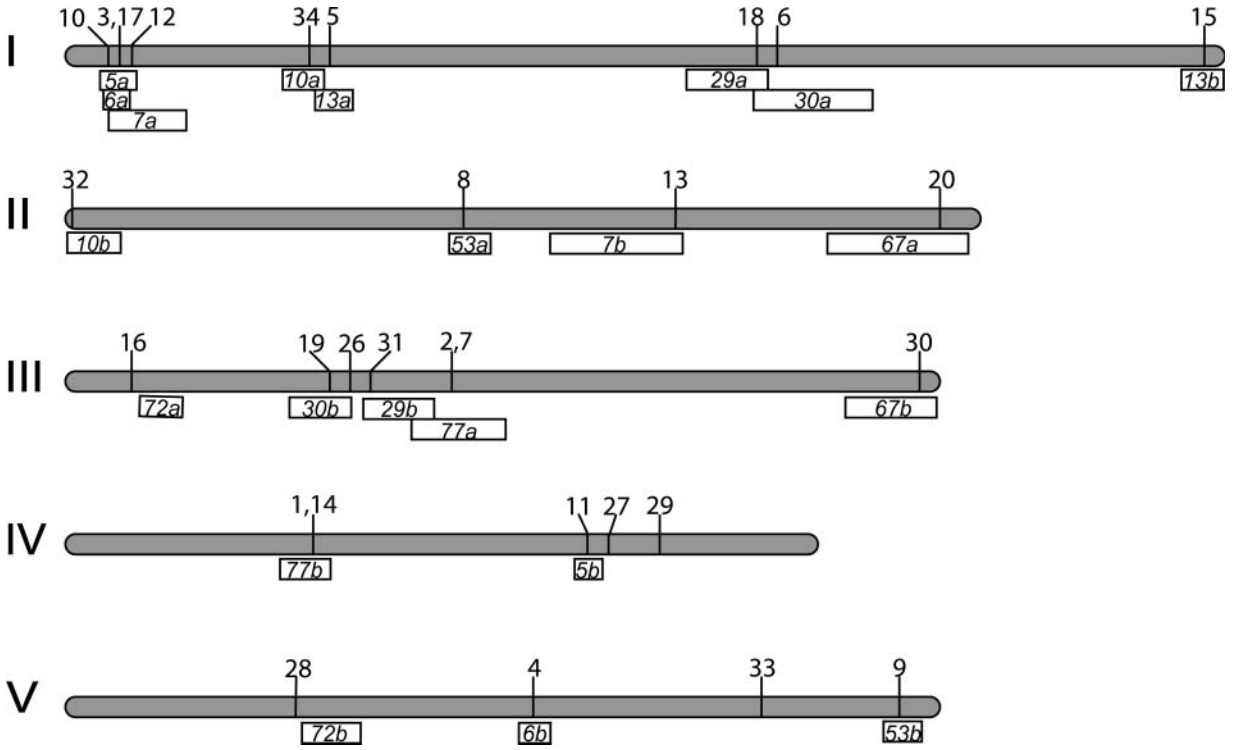

Figure 3. Locations of *Aux/IAA* loci with respect to duplicated chromosomal blocks per Vision et al. (2000). Numbers above chromosomes (shaded bars) refer to the *IAA* locus number. Numbered white bars below chromosomes indicate locations of duplicated segments identified in Vision et al. (2000).

duplications. At least 20 losses of individual duplicated loci during the *Aux/IAA* evolutionary history are also required. Node *c* is inferred to be a tandem duplication in our reconstruction methodology (see "Materials and Methods" and supplemental material, which can be viewed at www.plantphysiol.org) because the only informative pair of loci for this event (*IAA18* and *IAA31* in block 29) was not identified as anchor loci for the block 29 duplication (Vision et al., 2000; Blanc et al., 2003). We also treated node *e* as a tandem duplication. The informative pair of loci for this event (*IAA5* and *IAA15* in block 13) was identified as possible anchor loci (Vision et al., 2000; Blanc et al., 2003), but an analysis of aligned nucleotide sequences for these two genes indicated that the level of synonymous substitutions is nearly double that of any of the sister locus pairs (data not shown). Consequently, we considered it more likely that the ancestors of *IAA5* and *IAA15* were neighboring genes on the ancestral segment of block 13 and arose from an earlier tandem duplication.

We also evaluated an alternate scenario, in which the *Aux/IAA* tree is rerooted such that the group B sequences are treated as monophyletic, as described by Rogg et al. (2001). This alternate reconstruction also involves 20 separate duplication events but only 18 losses of duplicated genes (Supplemental Figs. 1 and 2). Additional scenarios involving some changes to weakly supported branches in the phylogenetic tree topology require as few as nine gene losses (data not shown). All of these reconstructions, however, still require multiple rounds of tandem duplication as the initial steps in the proliferation of the nonbasal *Aux/IAA* loci, followed by multiple block duplications.

We used the base reconstruction to evaluate the proportion of nontandemly duplicated *Aux/IAA* loci in which both duplicated loci have been retained (Fig. 4). At least 24 segmental duplications of *Aux/IAA* loci involving blocks identified by Vision et al. (2000) must have occurred, including instances in which loci are contained in more than one block. In 12 of these duplications (50%), *Aux/IAA* loci are represented in both homologous segments. Under the base reconstruction, however, two of these pairs of loci are nonhomeologous, yielding a modified estimate of 26 segmental duplications with retention of both duplicates in ten cases (38%). When all inferred duplication events are considered, the base reconstruction depicts 39 nontandem gene duplications, with both duplicates retained until the next duplication event or until the present in 18 cases (46%). Only two of these gene duplications (represented by nodes *h* and *l*) entail duplication events that involve a single inferred ancestral locus and are considered to represent dispersed duplications. The remaining nontandem duplications each involve two or more neighboring loci and so represent segmental duplications. Several of these duplication events (those containing nodes *f*, *g*, *m*, *n*,

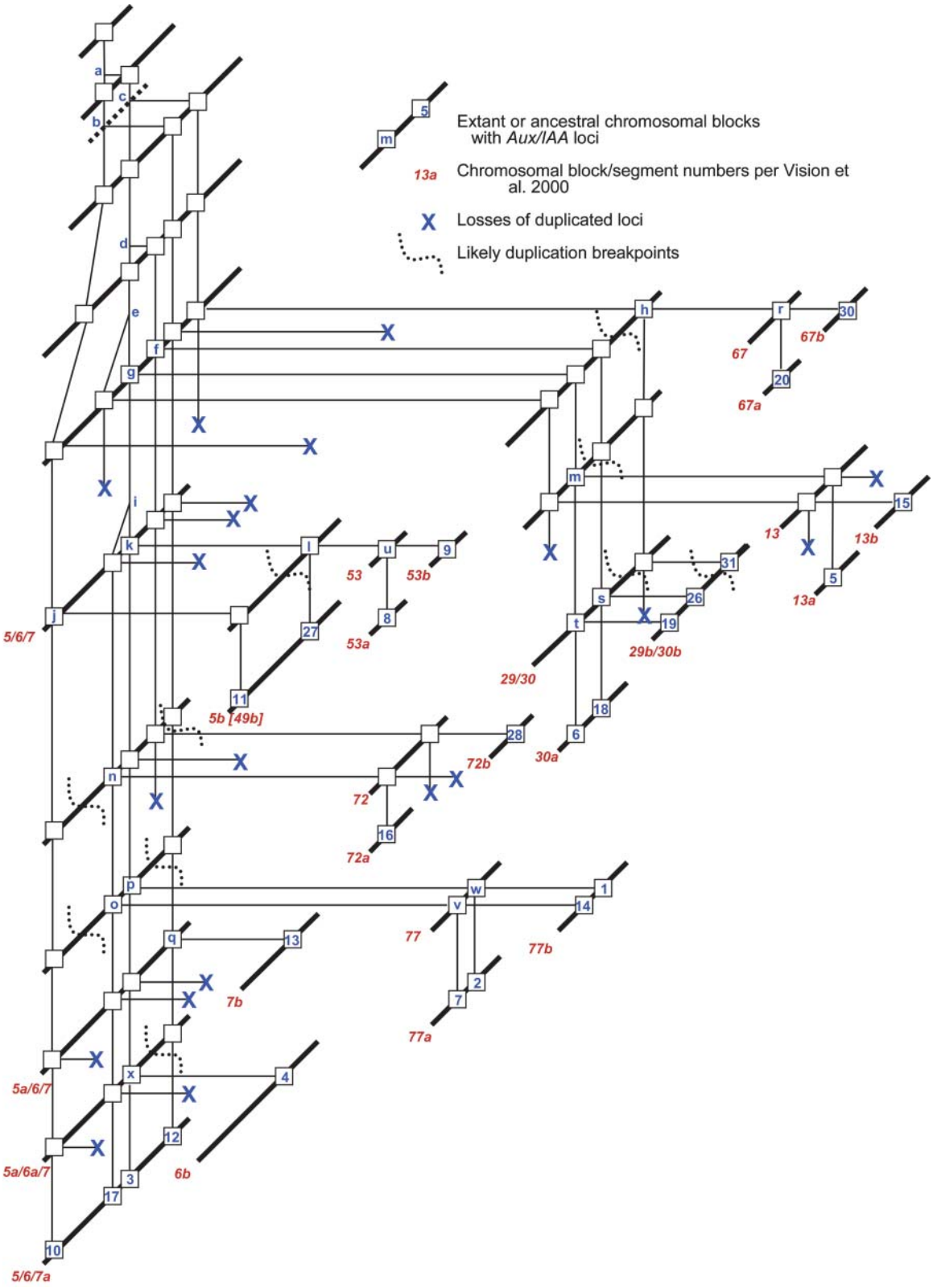

Figure 4. A hypothetical reconstruction of the evolutionary history of the *Aux/IAA* family (excluding *IAA29, IAA32, and IAA34*) in Arabidopsis. Line segments connecting loci on different chromosomal segments track the history of gene duplications, with a topology identical to that in Figure 1. Letter designations of nodes correspond to those in Figure 1.

o, and *p*) were not identified by Vision et al. (2000) but are directly or indirectly suggested in the more recent analysis of Blanc et al. (2003).

Relationship of *ARF* Phylogeny to Chromosomal Duplications

We used a similar approach to reconstruct the phylogeny of 23 Arabidopsis *ARF* sequences and evaluate their association with duplicate chromosomal blocks (Fig. 5). In contrast with the *Aux/IAA* family, only one out of eight *ARF* sister locus pairs was located in homologous segments. Retained duplicate *ARF*s were present in only three of 22 blocks containing *ARF*s (14%). Seven of the eight class I′ *ARF* loci comprising a single cluster (Fig. 5, node *o*) are located near each other in a region proximal to the chromosome 1 centromere and appear to be the products of a recent series of tandem duplications (Hagen and Guilfoyle, 2002). The relative branch lengths indicate that this cluster has evolved more rapidly than the remainder of the *ARF* family. One of the loci (*ARF 23*) contains a premature stop codon, indicating that it is probably a pseudogene.

Phylogenetic Relationships of Arabidopsis, *Medicago truncatula*, Rice, and Bryophyte *Aux/IAA* Sequences

In order to evaluate the divergence dates among the Arabidopsis *Aux/IAA* loci, we expanded the neighbor-joining analysis to include *Aux/IAA* sequences from other taxa: 15 *Aux/IAA* sequences from the legume *M. truncatula*, 12 sequences from rice, one from the bryophyte *Physcomitrella patens* (Imaizumi et al.,

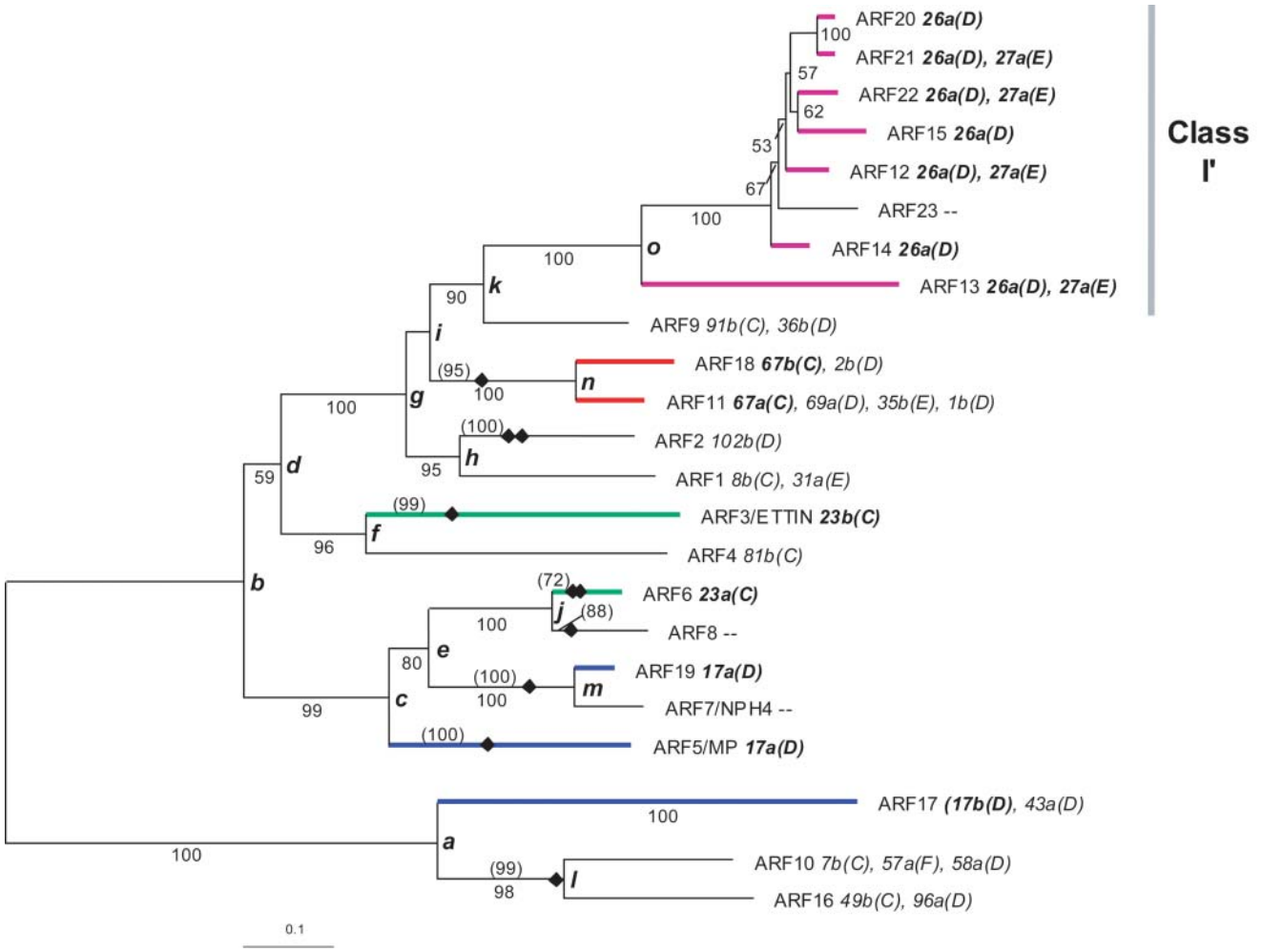

Figure 5. Neighbor-joining tree of Arabidopsis *ARF* loci. The percent bootstrap support for 500 replicates is shown below each branch with >50% support. All duplicated blocks per Vision et al. (2000), in which each sequence occurs, and the estimated age class (in parentheses) are listed after each locus name and are in bold if another *ARF* sequence is present in the block. Branches from the same putative chromosomal or local duplication are shown in the same colors. Diamonds denote positions of branches leading to one or more rice *ARF* loci. Numbers in parentheses above branches indicate percent bootstap support for rice *ARF* positions with >50% support.

2002), and a sequence from *Pinus pinaster*. In three phylogenetic reconstructions from alternative alignments (Fig. 6), *M. truncatula* sequences, either singly or in pairs, were unambiguously resolved as sister to four of the Arabidopsis *Aux/IAA* sister pairs, and occurred in more ambiguous positions basal to three additional Arabidopsis sister pairs. No *M. truncatula* loci were sister to individual Arabidopsis loci among the 10 sister pairs. These results provide evidence that the most recent round of Arabidopsis chromosomal duplication occurred after the eurosids I/II divergence separated the Arabidopsis and *M. truncatula* lineages approximately 96 to 113 Mya (Wikstrom et al., 2001), consistent with a number of recent studies (Blanc et al., 2003; Bowers et al., 2003; Raes et al., 2003).

Placement of most of the rice *Aux/IAA* sequences was ambiguous, due in part to the poor resolution of the main Arabidopsis group A subgroups. However, rice sequences occurred in consistent positions sister to *IAA7-14-17* and *IAA18-26*, and *IAA31*. Paradoxically, one rice sequence (OsTIGR7) was sister to *IAA26/PAP1*, a member of an Arabidopsis sister pair, but OsTIGR7 appears to be a partial sequence containing only the motif III-IV region, and its placement may be an artifact of estimating divergence from partial sequence data. Two of the *M. truncatula* sequences (MtTC38883 and MtNF051G11) also appear to be incomplete, which may affect the accuracy of their placement in the phylogenetic reconstruction. These results do suggest that the major subgroups within the group A and group B sequences, and even the divergence of *IAA16* from the *IAA7-14-17* cluster and the divergence of *IAA28* from *IAA18-26*, occurred before the divergence of the two lineages leading to Arabidopsis and rice 136 to 168 Mya (Wikstrom et al., 2001). A few of the date ranges for Arabidopsis chromosomal duplications implied by these data conflict with the initial estimates of Vision et al. (2000; see "Discussion"). The *P. patens* sequence is well supported as part of the cluster containing the group A genes, but basal to group A loci themselves. The very existence of an *Aux/IAA* locus in this bryophyte shows that the *Aux/IAA* family dates back at least to the origin of land plants, and its position in the tree suggests that the

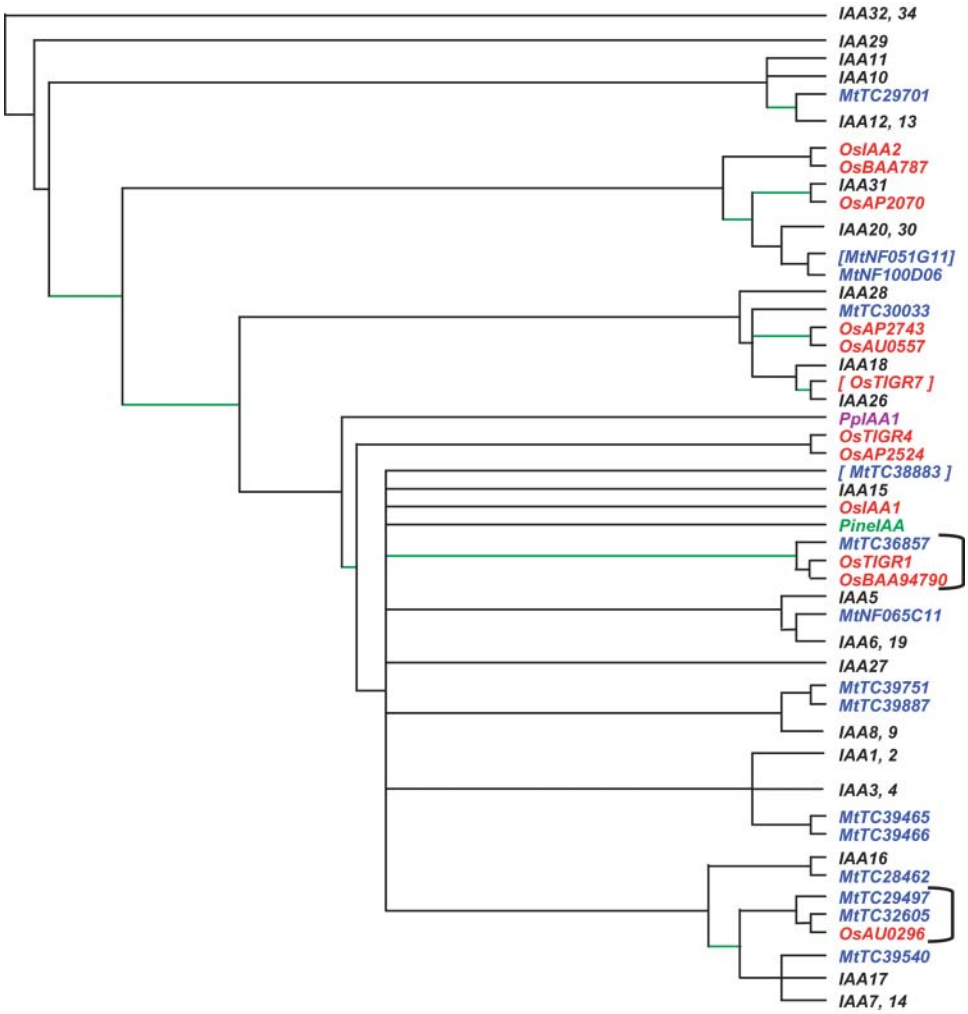

Figure 6. Consensus of three neighbor-joining trees of *Aux/IAA* loci from Arabidopsis, *M. truncatula*, rice, *P. patens*, and *P. pinaster*, derived from three alternate alignments of less-conserved regions. Branches shown in green had less than 50% bootstrap support in at least one of the three reconstructions. Clusters consisting only of Arabidopsis loci are shown as a single branch. Half brackets denote clusters of non-Arabidopsis loci whose placement in all three trees indicates a lineage that has been lost in Arabidopsis. The *OsTIGR7*, *MtTC38883*, and *MtNF051G11* sequences (shown in square brackets) are substantially incomplete, which could affect their placement.

family may be much older. The *P. pinaster* sequence was nested within the group A sequences in all three reconstructions, but its placement was inconsistent and lacked bootstrap support.

Some of the *M. truncatula* and rice sequences occurred in more basal positions relative to those described above (Fig. 6) and appear to represent at least two additional *Aux/IAA* subgroups that lack Arabidopsis counterparts. These subgroups could correspond to lost lineages in the base reconstruction depicted in Figure 4.

When 14 rice *ARF* sequences were included in the phylogenetic analysis of *ARF* loci, all occurred in positions that suggested sister relationships to eight individual Arabidopsis *ARF* loci or to sister pairs (Fig. 5). One or more rice *ARF* loci were paired with individual Arabidopsis *ARF* loci in three separate sister pairs. This suggests that at least nodes *a* to *j* in the *ARF* phylogeny (Fig. 5) represent duplications that occurred prior to the monocot-eudicot divergence. An expressed sequence tag (EST) sequence from *P. patens* in the public databases (accession no. BQ827439) appears to encode part of an *ARF* DNA-binding domain. When a BLAST search of the public protein databases was done using BQ827439 as query, the strongest matches were to ARF proteins. However, the sequence fragment was too short to include in our phylogenetic analyses.

DISCUSSION

Phylogenetic Relationships Among *Aux/IAA* and *ARF* Loci

Our analysis provides a comprehensive phylogenetic reconstruction of the *Aux/IAA* and *ARF* gene families. By using a set of *ARF* sequences as an outgroup, we identified the group B *Aux/IAA* sequences as a nested set of subgroups basal to a monophyletic group A, rather than a monophyletic sister clade to group A, as was supposed previously (Rogg et al., 2001). The apparent paraphyly of group B requires either that the *ARF* and *Aux/IAA* loci are sister gene families or that the *Aux/IAA* family arose from an ancestral *ARF* locus. Alternatively, the *ARF* family could have originated from an ancestral *Aux/IAA* locus via substitution of N-terminal regions. The N-terminal DNA-binding domain in ARF proteins is homologous to the B3 DNA-binding domain found in other families of plant proteins (Hagen and Guilfoyle, 2002). Under this scenario, the *P. patens Aux/IAA* sequence could conceivably be basal to all the *Aux/IAA* and *ARF* loci, resulting in a monophyletic group consisting of the group B *Aux/IAA* loci plus the *ARF* family, *IAA29*, *IAA32*, and *IAA34*, that is sister to the group A loci.

The evidence that *P. patens* also contains an *ARF* locus, however, makes monophyly of group B unlikely. The existence of a *P. patens ARF* locus, combined with the position of the *P. patens Aux/IAA* sequence, requires that the *IAA-ARF* divergence must at least predate the origin of group A (Fig. 1, node *g*). Consequently, the alternate and slightly more parsimonious *Aux/IAA* evolutionary history scenario that is possible with a monophyletic group B (Supplemental Figs. 1 and 2) also appears to be unlikely. Parsimony alone is not a reliable criterion for reconstruction of gene duplication histories (Gu and Huang, 2002) due to the high rate at which duplicated genes can be individually lost (Lynch and Conery, 2000; Wolfe, 2001). For example, parsimony criteria have been shown to favor an almost certainly incorrect model of individual gene duplications and translocations, rather than segmental duplications, to explain the patterns observed in the Arabidopsis genome (Gu and Huang, 2002). Reconciling chromosomal history and gene-family phylogeny, as we have done, minimizes the extent to which the history of gene loss is oversimplified. While our phylogenetic reconstruction itself may be incorrect in some of its details, two lines of evidence suggest that it is at least realistic. First, the base reconstruction (Fig. 4) requires 20 losses of duplicate genes in 39 nontandem gene duplications (representing 15 inferred or previously identified blocks). At least 12 of 24 *Aux/IAA* gene duplication events associated with blocks involve losses of at least one duplicate, so the overall predicted rate of duplicate gene loss is not excessive relative to the 85% to 91% loss rate seen in these blocks genome wide (Vision et al., 2000). Secondly, the occurrence of rice and *M. truncatula Aux/IAA* loci in subgroups that lack Arabidopsis sequences also indicates the loss of ancestral *Aux/IAA* genes in the Arabidopsis lineage. These losses are consistent with estimates that 10% to 15% of genes present in other Rosid and Asterid eudicots are absent from Arabidopsis (Allen, 2002).

Reasonable alternatives to some of the assumptions in our evolutionary history reconstruction methodology can be envisioned. Alternate phylogenetic reconstructions would lead to different evolutionary histories, a consideration we did explore to some extent. It is also possible that some loci are located near each other within existing segments or in homologous segments by coincidence due to chromosomal rearrangements rather than tandem and segmental duplications. The conserved colinearity of gene order that was used to identify the chromosomal blocks in the first place (Vision et al., 2000) and confirmation from a subsequent analysis of chromosomal duplications (Blanc et al., 2003), however, provide evidence that our methodology has produced a realistic evolutionary history scenario (see below).

Ages of *Aux/IAA* Family and Chromosomal Duplications

The presence of at least one *Aux/IAA* gene in Physcomitrella indicates that the *Aux/IAA* family dates to near the time of origin for land plants. The *P. patens Aux/IAA* gene has been found to be auxin-regulated, indicating that aspects of *Aux/IAA* function have also

been conserved in land plants (Imaizumi et al., 2002). Under the likely scenario that the *ARF* and *Aux/IAA* sequences comprise sibling gene families, the *Aux/IAA* family would have already undergone a number of duplications by the time the bryophyte and vascular plant lineages diverged some 450 to 700 Mya (Hedges, 2002). All of the major subgroups of group A and B *Aux/IAA* loci and most ancestors of *ARF* sister pairs appear to have originated before the monocot-eudicot divergence 136 to 168 Mya (Wikstrom et al., 2001). Most of the sister pairs of Arabidopsis *Aux/IAA* sequences, which appear to have originated during the most recent round of genomic duplication in Arabidopsis, arose after the divergence of the eurosids I and II clades.

Using a tentative protein sequence divergence clock, Vision et al. (2000) estimated the date of the most recent duplication of the Arabidopsis genome (age class C) at approximately 100 Mya. The topology of *M. truncatula* loci relative to Arabidopsis sister pairs is consistent with this estimate if the oldest of the estimated dates for the eurosid I-II divergence is used. More recent estimates have placed this duplication event within the eurosids II clade, substantially after the divergence of the lineages leading to Arabidopsis and *Gossypium hirtum* (Blanc et al., 2003). That split is estimated to have occurred 81 to 94 Mya (Wikstrom et al., 2001).

Four of the blocks involving Arabidopsis sister pairs were assigned to age class D by Vision et al. (2000), but the positions of *M. truncatula Aux/IAA* loci indicate that at least block 30 and either block 6 or block 77 diverged more recently, within the eurosids II lineage. Also, the block 6 and 7 duplications must have been separate events if these two blocks do in fact overlap. The base evolutionary history reconstruction (which assumes that *IAA27/PAP2* also belongs to segment 5b) also favors an earlier origin for block 5. The results reported here are consistent with the findings of Blanc et al. (2003), who conclude that blocks 6, 30, and 77 are recent and contemporaneous with each other, while block 5 is older, and Raes et al. (2003), who dated blocks 6 and 30 as being approximately 70 My old and block 5 as approximately 135 My old. Another discrepancy is block 53, which was assigned to age class C by Vision et al. (2000), but which Blanc et al. (2003) propose belongs to an earlier age class.

All our evolutionary history reconstructions, including those in which the phylogenetic tree constraint was relaxed, suggest that nearly all of the early branching points in the *Aux/IAA* phylogeny were tandem duplications. The apparent antiquity of the gene family requires that the initial duplications must have occurred near or before the emergence of land plants. The 5a-6a-7a region of chromosome 1, which contains four *Aux/IAA* loci in separate sublineages, may resemble the arrangement of the ancestral Aux/IAA genes. One intriguing implication of this hypothesis is that this region represents an intact remnant of the ancestral land plant genome that has not been broken up by chromosomal rearrangements for perhaps hundreds of millions of years, a possibility anticipated by Paterson et al. (1996).

Evidence for Predicted Older Duplications

In addition to the block duplications identified by Vision et al. (2000), a number of additional nontandem gene duplications are inferred in the base reconstruction of the *Aux/IAA* family. Four additional inferred duplication events involve multiple ancestral loci, suggesting that they may represent older duplicated blocks rather than dispersed duplications of individual genes. At least three of these older duplications are also supported by a more recent analysis (Blanc et al., 2003). The ancestor of block 77 and at least part of blocks 5 and 6 appears to represent an older duplicated block, attested by nodes *o* and *p* (Figs. 1 and 4) and by Blanc et al. (2003). Secondly, block 72 also appears to share a common origin with at least some of blocks 5 and 6, represented by node *n*, assuming that *IAA16* and/or *IAA28* are associated with block 72 but have lost their respective duplicates. Blanc et al. (2003) include *IAA16*, but not *IAA28*, within the block 72 region and identify node *n* as part of an old segmental duplication. Thirdly, blocks 13 and 30 appear to share a common ancestor that includes node *m*, which is also verified by Blanc et al. (2003). Finally, an early duplication of a chromosomal segment containing the ancestors of all modern group A and B loci is putatively represented by nodes *f* and *g*. The segment 5a-6a-7a and 29a-30a regions, respectively, appear to be the most extensive intact remnants of this inferred block. Blanc et al. (2003) associate node *f* with a segmental duplication, but our association of this node with the more extensive duplication of *Aux/IAA* loci that also includes node *g* depends on the assumption that *IAA28* is actually part of segment 72b.

The Blanc et al. (2003) analysis also verifies that nodes *j* and *k* belong to the same duplication event, as our reconstruction predicts, with *IAA27* included in segment *5b*. Their analysis also provides evidence of the ancient tandem duplication of node *b* and that this duplication involved a multiple-gene region corresponding to our predicted tandem duplication involving nodes *b* and *c*. Overall, the Blanc et al. (2003) study provides extensive confirmation for the major features of our evolutionary history reconstruction. Many of the details remain uncertain, however, and the alternate scenario (Supplemental Fig. 2) is also largely consistent with the Blanc et al. (2003) analysis.

Evolutionary and Functional Implications

One of the most striking findings of this study is the correspondence of all 10 *Aux/IAA* sister locus pairs with block duplications and an overall elevated level of retention for segmentally duplicated *Aux/IAA* genes. Throughout the genome, only about 15% of

dispersed (i.e. nontandem) duplicated gene pairs have been found to be associated with duplicated chromosomal blocks (Vision et al., 2000). Thus, it is remarkable that all of the most recent duplication events represented in the extant *Aux/IAA* family are associated with such blocks.

By contrast, relatively little diversification has occurred in the *ARF* family since the monocot-dicot divergence except for the recent tandem proliferation that produced the class I′ *ARF* subgroup, and at least some of the eight loci in this cluster are likely to be pseudogenes. The recent episode of tandem proliferation in the *ARF* family is another interesting contrast with the *Aux/IAA* family, in which there is no evidence of tandem duplication events within the last approximately 100 Myr. Only one of eight *ARF* sister pairs was associated with a duplicated block, a ratio more typical of the Arabidopsis genome. It appears that the most recent round of genomic duplication within the Eurosids II lineage, which may have given rise to nearly all of the *Aux/IAA* sister pairs, produced almost no long-term expansion of the *ARF* family. While doubtless many duplicated blocks remain to be identified, it would be surprising if the *ARF* genes were, as a group, to be preferentially represented in unidentified blocks. However, we cannot exclude the possibility that the early branching events in the *ARF* family were the result of extremely ancient segmental duplications that are now undetectable.

Why, then, have so many of the segmental duplications of *IAA* genes persisted? One hypothesis is that *Aux/IAA* loci that are duplicated simultaneously with the rest of the genome might be more viable than those duplicated singly, as they would then maintain proper dosage relationships with interacting proteins. Aux/IAA proteins regulate gene expression indirectly by interacting with ARF proteins and with auxin signaling mechanisms. Aux/IAA proteins can dimerize with each other as well as forming IAA/ARF heterodimers (Kim et al., 1997), so degenerative mutations in duplicated genes could deleteriously affect the normal function of these complexes (Hughes and Hughes, 1993; Gottlieb and Ford, 1997). Segmentally duplicated genes encoding 20S proteosome subunits also appear to have been preferentially retained in Arabidopsis, suggesting that loss of stoichiometry is costly for these multimeric protein complexes (Cannon and Young, 2003). It is possible that dimerizing Aux/IAA proteins inhibit each other's activities in order to maintain appropriate regulatory homeostasis. Consistent with this idea, gain-of-function mutations in several *IAA* genes cause contrasting phenotypes. For example, gain-of-function *iaa14/slr*, *iaa3/shy2*, and *iaa28* mutants, which represent both major *Aux/IAA* groups and two distinct group A subgroups, have reduced numbers of lateral roots (Tian and Reed, 1999; Rogg et al., 2001; Fukaki et al., 2002), whereas gain-of-function *iaa7/axr2* and *iaa17/axr3* mutants, from the same subgroup as *IAA14/SLR*, show the opposite phenotype (Liscum and Reed, 2002). Moreover, IAA3/SHY2 and IAA17/AXR3, encoded by adjacent genes, have been shown to interact antagonistically to regulate root hair development (Knox et al., 2003). Such balancing need not act solely on proteins that interact physically, as different Aux/IAA proteins might instead act in different tissues to maintain proportional auxin responses in different cell types or organs. The quite distinct expression patterns of $P_{SHY2/IAA3}$::*GUS* (β-glucuronidase) and $P_{AXR2/IAA7}$::*GUS* (Tian et al., 2002) suggest that a more indirect model of this type is plausible. By contrast, the products of *ARF* genes directly regulate transcription as DNA-binding proteins (Hagen and Guilfoyle, 2002). This more direct regulatory mechanism may have resulted in minimal constraints on the degenerative loss of duplicated sets of *ARF* genes.

The hypothesis described above does not preclude the possibility that some of the retained segmental duplicates may have undergone subsequent divergence in function, either through subtle changes in their interactions with other proteins or in their expression patterns. Some lines of evidence support a degree of functional divergence between sister segmental duplicate *Aux/IAA* genes. Mutants at sister segmental duplicates *IAA7/AXR2* and *IAA14/SLR* display contrasting root development phenotypes, as discussed above. *IAA7/AXR2* and *IAA8* both require de novo protein synthesis for auxin-responsive expression, but their respective sister loci do not (Abel et al., 1995). However, evolution of new developmental roles (Ohno, 1970) or complementary loss of multiple ancestral functions (Force et al., 1999) do not explain why segmentally duplicated *Aux/IAA* genes would have been preferentially maintained over individually duplicated loci.

Another possibility is that remote cis-regulatory elements required for *Aux/IAA* transcription are retained only when sufficiently large chromosomal regions are duplicated. Under this model, more localized duplications of chromosomal segments containing *Aux/IAA* genes without regulatory elements would result in nonfunctional genes. Remote enhancers have been found to regulate expression of several mammalian regulatory genes, including *HoxD* cluster genes (Herault et al., 1997; Kmita et al., 2002), β-globin genes (Dillon et al., 1997), and *Sonic hedgehog* (Lettice et al., 2002), and long-distance regulatory elements have also been found to be required for paramutation at the maize (*Zea mays*) *b1* locus (Stam et al., 2002). In the first two of these cases, regulation occurs in a distance-dependent manner that affects the relationship between locus order and expression patterns. An analogous mechanism in the *Aux/IAA* family could explain both the preferential preservation of loci in duplicate blocks and differences in mutant phenotypes among *Aux/IAA* genes.

Our results should provide useful guidance for further *Aux/IAA* functional studies. In particular, possible functions of protein regions outside the four conserved motifs should be considered. We observed short regions with considerable protein sequence

Table I. *Sources of Aux/IAA and ARF sequences used in this study*

Locus	Data Source[a]	Location or Database Reference	Type of Sequence[b]
Arabidopsis *Aux/IAA* Loci			
IAA1	TAIR	At4g14560	
IAA2	TAIR	At3g23030	
IAA3/SHY2	TAIR	At1g04240	
IAA4	TAIR	At5g43700	
IAA5	TAIR	At1g15580	
IAA6	TAIR	At1g52830	
IAA7/AXR2	TAIR	At3g23050	
IAA8	TAIR	At2g22670	
IAA9	TAIR	At5g65670	
IAA10	TAIR	At1g04100	
IAA11	TAIR	At4g28640	
IAA12/BDL	TAIR	At1g04550	
IAA13	TAIR	At2g33310	
IAA14/SLR	TAIR	At4g14550	
IAA15	TAIR	At1g80390	Predicted protein
IAA16	TAIR	At3g04730	
IAA17/AXR3	TAIR	At1g04250	
IAA18	TAIR	At1g51950	
IAA19/MSG2	TAIR	At3g15540	
IAA20	TAIR	At2g46990	
IAA26/PAP1	TAIR	At3g16500	
IAA27/PAP2	TAIR	At4g29080	
IAA28	TAIR	At5g25890	
IAA29	TAIR	At4g32280	[c]
IAA30	TAIR	At3g62100	
IAA31	TAIR	At3g17600	
IAA32	TAIR	At2g01200	Partial EST only
IAA33	TAIR	At5g57420	Predicted protein
IAA34	TAIR	At1g15050	[c]
M. truncatula Aux/IAA Loci			
MtTC39540	TIGR	TC39540 [d]	cDNA
MtTC39465	TIGR	TC39465 [d]	cDNA
MtTC39466	TIGR	TC39466 [d]	cDNA
MtTC28462	TIGR	TC28462	cDNA
MtTC29497	TIGR	TC29497 [d]	cDNA
MtTC32605	TIGR	TC32605 [d]	cDNA
MtTC38883	TIGR	TC38883 [d]	cDNA[c]
MtTC39887	TIGR	TC39887 [d]	cDNA[c]
MtTC39751	TIGR	TC39751 [d]	cDNA[c]
MtTC36857	TIGR	TC36857 [d]	cDNA[c]
MtTC29701	TIGR	TC29701 [d]	cDNA[c]
MtTC30033	TIGR	TC30033 [d]	cDNA
MtNF065C11	TIGR	NF065C11EC1F1085	EST[c]
MtNF051G11	TIGR	NF051G11EC1F1086	EST
MtNF100D06	TIGR	NF100D06EC1F1048	EST
Rice *Aux/IAA* Loci			
OsAP2070	GenBank	BAA95840	
OsBAA78739	GenBank	BAA78739	
OsAP2743	GenBank	BAA99424	
OsAU055784	GenBank	AU055784	EST
OsAP2524	GenBank	BAB07974	
OsAU029620	GenBank	AU029620	EST
OsBAA94790	GenBank	BAA94790	
OsIAA1	GenBank	CAC80823	cDNA
OsIAA2	GenBank	AC069158_20	Predicted protein
OsTIGR1	TIGR	TC56753 [d]	cDNA
OsTIGR4	TIGR	TC63581 [d]	cDNA
OsTIGR7	TIGR	TC60579 [d]	cDNA

(*Table continues on following page.*)

Table I. (*Continued from previous page.*)

Locus	Data Source[a]	Location or Database Reference	Type of Sequence[b]
P. pinaster Aux/IAA Locus			
PineIAA	GenBank	CAC85936	cDNA
P. patens Aux/IAA Locus			
PpIAA1	GenBank	BAB71766	cDNA
Arabidopsis *ARF* Loci			
ARF1	TAIR	At1g59750	
ARF2	TAIR	At5g62010	
ARF3/ETTIN	TAIR	At2g33860	
ARF4	TAIR	At5g60450	
ARF5/MP	TAIR	At1g19850	
ARF6	TAIR	At1g30330	
ARF7/NPH4	TAIR	At5g20730	
ARF8	TAIR	At5g37020	
ARF9	TAIR	At4g23980	
ARF10	TAIR	At2g28350	
ARF11	TAIR	At2g46530	Partial EST only[c]
ARF12	TAIR	At1g34310	Predicted protein
ARF13	TAIR	At1g34170	Predicted protein[c]
ARF14	TAIR	At1g35540	Predicted protein
ARF15	TAIR	At1g35520	Predicted protein[c]
ARF16	TAIR	At4g30080	[c]
ARF17	TAIR	At1g77850	[c]
ARF18	TAIR	At3g61830	
ARF19	TAIR	At1g19220	
ARF20	TAIR	At1g35240	Predicted protein[c]
ARF21	TAIR	At1g34410	Predicted protein
ARF22	TAIR	At1g34390	Predicted protein
ARF23	TAIR	At1g43950	Predicted protein[c]
Rice *ARF* Loci			
OsAC024594	GenBank	AC024594_6	
OsBAB89547	GenBank	BAB89547	
OsARF1	GenBank	AF140228_1	cDNA
OsARF2	GenBank	BAB85913	cDNA
OsETTIN-like1	GenBank	BAB85910	cDNA
OsETTIN-like2	GenBank	BAB85911	cDNA
OsMP-like	GenBank	BAB85912	cDNA
OsARF6a	GenBank	BAB85914	cDNA
OsARF6b	GenBank	BAB85915	cDNA
OsARF7a	GenBank	BAB85916	cDNA
OsARF7b	GenBank	BAB85917	cDNA
OsARF8	GenBank	BAB85918	cDNA
OsARF10	GenBank	BAB85919	cDNA
OsARF16	GenBank	BAB85920	cDNA

[a]Data sources: TAIR, www.arabidopsis.org; GenBank, www.ncbi.nlm.nih.gov; TIGR, www.tigr.org. [b]Sequences are translations from annotated genomic DNA sequences with cDNA verification of expression unless otherwise noted in this column. [c]Database versions of sequences edited by the authors. [d]Assembly also includes other related sequences.

similarity between group A subgroups, especially in the regions to the N-terminal side of motif I, between motifs I and II, and between motifs III and IV. Considering that the group A subgroups appear to have diverged more than 150 Mya, this sequence conservation implies substantial selective constraint on these regions. It would be of interest to determine whether mutations occurring outside the conserved motifs I to IV have visible phenotypes that would help identify possible functional roles.

Sequence conservation is also observed among several *Aux/IAA* sister pairs in upstream flanking sequences containing shared AuxRE motifs (Fig. 2). These AuxREs are likely to be functionally important in the observed transcriptional activation of *Aux/IAA* genes by auxin (Gray et al., 2001; Tiwari et al., 2001, 2003; Hagen and Guilfoyle, 2002). Consequently, conservation of these motifs and surrounding regions suggests that transcriptional regulatory patterns are likely to have been conserved among sister loci as well.

The ages of the *Aux/IAA* lineages and relative branch lengths in the neighbor-joining tree (Fig. 1) also argue that all of the genes are likely to be functional. Transcripts of *IAA15* could not be detected by northern hybridization (Abel et al., 1995), leading to the suggestion that it may be a pseudogene. None of the Arabidopsis *Aux/IAA* sequences, however, have premature stop codons or inordinately long branch lengths that would be characteristic of pseudogenes. By contrast, the *ARF* subgroup consisting of the loci from the recent tandem proliferation plus *ARF23* appears more likely to contain one or more pseudogenes. This subgroup is separated from its sister locus, *ARF9*, by a long internal branch indicating accelerated evolution due to a possible loss of selective constraint, and *ARF23* contains an internal stop codon. A search of public databases revealed no ESTs that would provide evidence of expression for any of the genes in this subgroup, with the exception of a single EST similar to *ARF14* (Hagen and Guilfoyle, 2002).

SUMMARY

Studies of chromosomal duplications in Arabidopsis have already proved useful for phylogenetic analyses of gene families and vice versa (Barakat et al., 2001; Rosenquist et al., 2001; Vandepoele et al., 2002). In this study, we have combined traditional phylogenetic analysis with information on chromosomal duplications in Arabidopsis to obtain insight into both genome evolution and the biology of the *Aux/IAA* and *ARF* gene families. One useful outcome has been to obtain refined estimates for the ages of several chromosomal block duplications relative to the divergence of major angiosperm lineages. This approach shows great promise in allowing a more detailed reconstruction of the evolutionary history of plant genomes than would be possible in the absence of phylogenetic information. Secondly, we have identified possible additional duplications not detected in the earlier analysis (Vision et al., 2000), using reasoning similar to that of other recent studies (Simillion et al., 2002; Blanc et al., 2003). These additional duplications were found to be largely consistent with the results of the analysis by Blanc et al. (2003). Finally, we obtained evidence for biased preservation of duplicated *Aux/IAA* loci, but not *ARF* loci, in chromosomal blocks within the Arabidopsis lineage, which raises new questions about the modes of diversification in these two gene families.

MATERIALS AND METHODS

Sequence Data, Alignments, and Phylogenetic Reconstructions

Experimentally determined or predicted amino acid sequence data for Arabidopsis, rice (*Oryza sativa*), *Physcomitrella patens*, and *Pinus pinaster* Aux/IAA proteins available as of February 2002 were obtained from GenBank. *Medicago truncatula Aux/IAA* nucleotide sequences and translations and additional rice *Aux/IAA* sequences were obtained from The Institute for Genomic Research (TIGR; Rockville, MD; www.tigr.org) in October 2001 and February 2002, respectively, and were further edited manually to correct obvious frameshift errors in base calling or remove low quality sequence. Sources and accession numbers (where applicable) for all sequences are listed in Table I. Arabidopsis ARF amino acid sequences were also obtained from GenBank, and annotations were edited by the authors as noted in Table I. ARF protein sequences from rice were obtained from GenBank. Aux/IAA and ARF protein sequences were manually aligned. The primary alignments of translated *Aux/IAA* and *ARF* sequences used in this study are available from the authors at the following Web site: http://www.uncg.edu/~dlreming.

The Aux/IAA protein sequences could be aligned with a high degree of confidence in the conserved motifs I, II, III, and IV (Abel et al., 1995). Outside these motifs, alignments are reliable only between closely related sequences. Including these variable regions, however, provided useful resolution among more closely related sequences. This additional resolution was an important consideration with the Aux/IAA family, as the sequences are short (158–338 amino acids in Arabidopsis). Various possible alignments between dissimilar sequences appear to have comparable proportions of matching amino acids, so it is unlikely that alignment errors would greatly affect the results of distance-based phylogenetic analyses. To test the sensitivity of tree reconstruction to alignment ambiguities, we conducted analyses with three alternate alignments that differed in the more variable regions. An additional sequence (*IAA33*), which shows evidence of homology to Aux/IAA and ARF proteins but lacks most of motif III, was not included. Phylogenetic analysis of ARF protein sequences used only the conserved N-terminal DNA-binding domain and the conserved C-terminal region corresponding to the Aux/IAA motif III-IV region.

Neighbor-joining analyses of the Arabidopsis Aux/IAA and ARF sequences were conducted in PHYLIP 3.5 (Department of Genetics, University of Washington, Seattle; http://evolution.genetics.washington.edu/phylip.html) using the PAM matrix of Dayhoff (1979), with 500 bootstrap replicates and randomized sequence input order. Sites with gaps in pairwise comparisons were treated as missing data. Analyses including non-Arabidopsis Aux/IAA sequences were conducted in a similar manner, but only 100 bootstrap replicates were generated. Maximum parsimony analyses were also conducted using the PROTPARS algorithm of PHYLIP. Gaps were recoded so as to be treated as missing characters. In order to reciprocally root the Aux/IAA and ARF phylogenies, neighbor-joining and maximum parsimony trees were constructed from alignments of 71 sites in the homologous motif III-IV regions of the Aux/IAA proteins and seven ARF proteins representing the primary ARF subgroups (ARF2, -4, -5, -10, -11, -12, and -16).

Reconstruction of Gene Duplication Histories

Chromosomal positions of all known and predicted *Aux/IAA* loci were obtained from The Arabidopsis Information Resource (TAIR) database (http://www.arabidopsis.org/home.html). These were compared against the genomic duplication dataset of Vision et al. (2000), available at http://www.bio.unc.edu/faculty/vision/lab/arab/science_supplement, in order to identify duplicated blocks encompassing each locus. A block is defined as a pair of chromosome segments that are believed to be descended from a common ancestral segment (hereafter referred to as homologous segments). The chromosomal locations for some of the *Aux/IAA* genes were not listed in the duplication dataset, but cross-referencing with more recent assemblies of the genome allowed unambiguous determination of their locations with respect to blocks.

Reconstruction of the *Aux/IAA* evolutionary history involved a two-stage process (see supplemental material for details). In the first stage, each node of the phylogenetic tree was classified as a segmental, tandem, or dispersed duplication, starting with the most terminal nodes and working backward in topological order (Sedgewick, 1990). Classification of nodes was based on the occurrence and positions of loci in homologous duplicated chromosomal segments or their ancestral segments among the two daughter lineages of each node. Locus pairs in homologous segments were evaluated for their status as anchor loci for the inferred segmental duplication (Vision et al., 2000). The mode of duplication at some nodes could not be fully classified at this stage. In the second stage, duplication events involving ancestral chromosome segments were reconstructed in a forward direction, beginning with the inferred single ancestral locus. The reconstruction process resulted in a number of possible evolutionary history scenarios, which differ from each other in the order of some independent duplication events and in the mode of duplication at nodes that could not be fully classified in the first stage. We selected a single base scenario from among the various alternatives, based on additional

evidence, such as the relative sequence divergence of loci descending from common ancestors on the inferred ancestral chromosomal segments and the degree of support for putative anchor loci. We cannot ensure, however, that our methodology will identify all plausible evolutionary history scenarios, or that all the scenarios that it generates will be plausible.

ACKNOWLEDGMENTS

We thank Brandon Gaut and two anonymous reviewers for constructive suggestions on earlier versions of this manuscript.

Received January 23, 2004; returned for revision April 22, 2004; accepted April 26, 2004.

LITERATURE CITED

Abel S, Nguyen MD, Theologis A (1995) The *PS-IAA4/5*-like family of early auxin-inducible mRNAs in *Arabidopsis thaliana.* J Mol Biol **251:** 533–549

Allen KD (2002) Assaying gene content in *Arabidopsis.* Proc Natl Acad Sci USA **99:** 9568–9572

Barakat A, Szick-Miranda K, Chang I-F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J (2001) The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. Plant Physiol **127:** 398–415

Basu S, Sun H, Brian L, Quatrano RL, Muday GK (2002) Early embryo development in *Fucus distichus* is auxin sensitive. Plant Physiol **130:** 292–302

Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the arabidopsis genome. Plant Cell **12:** 1093–1101

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res **13:** 137–144

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Cannon SB, Young ND (2003) OrthoParaMap: distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. BMC Bioinformatics **4:** 35

Cooke TJ, Poli DB, Sztein AE, Cohen JD (2002) Evolutionary patterns in auxin action. Plant Mol Biol **49:** 319–338

Dayhoff MO (1979) Atlas of Protein Sequence and Structure, Vol 5, Suppl 3, 1978. National Biomedical Research Foundation, Washington, DC

Dillon N, Trimborn T, Strouboulis J, Fraser P, Grosveld F (1997) The effect of distance on long-range chromatin interactions. Mol Cell **1:** 131–139

Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545

Fukaki H, Tameda S, Masuda H, Tasaka M (2002) Lateral root formation is blocked by a gain-of-function mutation in the *SOLITARY-ROOT/IAA14* mutation of *Arabidopsis.* Plant J **29:** 153–168

Gottlieb LD, Ford VS (1997) A recently silenced, duplicate *PgiC* locus in *Clarkia.* Mol Biol Evol **14:** 125–132

Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M (2001) Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. Nature **414:** 271–276

Gu X, Huang W (2002) Testing the parsimony test of genome duplications: a counterexample. Genome Res **12:** 1–2

Hagen G, Guilfoyle TJ (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol Biol **49:** 373–385

Hedges SB (2002) The origin and evolution of model organisms. Nat Rev Genet **3:** 838–849

Herault Y, Fraudeau N, Zakany J, Duboule D (1997) *Ulnaless* (*Ul*), a regulatory mutation inducing both loss-of-function and gain-of-function of posterior *Hoxd* genes. Development **124:** 3493–3500

Hughes MK, Hughes AL (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis.* Mol Biol Evol **10:** 1360–1369

Imaizumi T, Kadota A, Hasebe M, Wada M (2002) Cryptochrome light signals control development to suppress auxin sensitivity in the moss *Physcomitrella patens.* Plant Cell **14:** 373–386

Kim J, Harter K, Theologis A (1997) Protein-protein interactions among the Aux/IAA proteins. Proc Natl Acad Sci USA **94:** 11786–11791

Kmita M, Fraudeau N, Herault Y, Duboule D (2002) Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. Nature **420:** 145–150

Knox K, Grierson CS, Leyser O (2003) AXR3 and SHY2 interact to regulate root hair development. Development **130:** 5769–5777

Ku H-M, Vision T, Liu J, Tanksley SD (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA **97:** 9121–9126

Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, et al (2002) Disruption of a long-range cis-acting regulator for *Shh* causes preaxial polydactyly. Proc Natl Acad Sci USA **99:** 7548–7553

Liscum E, Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol **49:** 387–400

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science **290:** 1151–1155

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Berlin

Paterson AH, Lan TH, Reischmann KP, Chang C, Lin YR, Liu SC, Burow MD, Kowalski SP, Katsar CS, DelMonte TA, et al (1996) Toward a unified genetic map of higher plants, transcending the monocot-dicot divergence. Nat Genet **14:** 380–382

Raes J, Vandepoele K, Saeys Y, Simillion C, Van de Peer Y (2003) Investigating ancient duplication events in the Arabidopsis genome. J Struct Funct Genomics **3:** 117–129

Reed JW (2001) Roles and activities of Aux/IAA proteins in *Arabidopsis.* Trends Plant Sci **6:** 420–425

Rogg LE, Lasswell J, Bartel B (2001) A gain-of-function mutation in *IAA28* supresses lateral root development. Plant Cell **13:** 465–480

Rosenquist M, Alsterfjord M, Larsson C, Sommarin M (2001) Data mining the Arabidopsis genome reveals fifteen 14-3-3 genes: expression is demonstrated for two out of five novel genes. Plant Physiol **127:** 142–149

Sedgewick R (1990) Algorithms in C. Addison-Wesley, Reading, MA

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana.* Proc Natl Acad Sci USA **99:** 13627–13632

Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL (2002) The regulatory regions required for *B′* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. Genetics **162:** 917–930

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815

Tian Q, Nagpal P, Reed JW (2003) Regulation of *Arabidopsis* SHY2/IAA3 protein turnover. Plant J **36:** 643–651

Tian Q, Reed JW (1999) Control of auxin-regulated root development by the *Arabidopsis thaliana SHY2/IAA3* gene. Development **126:** 711–721

Tian Q, Uhlir NJ, Reed JW (2002) Arabidopsis SHY2/IAA3 inhibits auxin-regulated gene expression. Plant Cell **14:** 301–319

Tiwari SB, Hagen G, Guilfoyle TJ (2003) The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell **15:** 533–543

Tiwari SB, Wang X-J, Hagen G, Guilfoyle TJ (2001) AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. Plant Cell **13:** 2809–2822

Ulmasov T, Hagen G, Guilfoyle TJ (1999a) Activation and repression of transcription by auxin-response factors. Proc Natl Acad Sci USA **96:** 5844–5849

Ulmasov T, Hagen G, Guilfoyle TJ (1999b) Dimerization and DNA binding of auxin response factors. Plant J **19:** 309–319

Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S, Inze D (2002) Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell **14:** 903–916

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis.* Science **290:** 2114–2117

Wikstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. Proc R Soc Lond B Biol Sci **268:** 2211–2220

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet **2:** 333–341