Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Cold Spring Harbor Symposia on Quantitative Biology

Evolutionary History of Plant Multisubunit RNA Polymerases IV and V : Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization

S.L. Tucker, J. Reece, T.S. Ream, et al.

Cold Spring Harb Symp Quant Biol published online March 29, 2011
Access the most recent version at doi:10.1101/sqb.2010.75.037

References This article cites 65 articles, 37 of which can be accessed free at:
http://symposium.cshlp.org/content/early/2011/03/21/sqb.2010.75.037.refs.html

P<P Published online March 29, 2011 in advance of the print volume.

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Advance online articles have not yet appeared in the print volume. Citations to Advance online articles must include the digital object identifier (DOI) and date of initial publication.

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:
http://symposium.cshlp.org/subscriptions

Copyright © 2011, Cold Spring Harbor Laboratory Press

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Evolutionary History of Plant Multisubunit RNA Polymerases IV and V

Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization

S.L. Tucker,[1] J. Reece,[1] T.S. Ream,[1,3] and C.S. Pikaard[1,2]

[1]*Biology Department, Washington University, St. Louis, Missouri 63130;* [2]*Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405*

Correspondence: cpikaard@indiana.edu

Eukaryotes have three multisubunit DNA-dependent RNA polymerases that are essential for viability, abbreviated as Pol I, Pol II, and Pol III. Remarkably, *Arabidopsis thaliana* and other higher plants contain two additional nuclear multisubunit RNA polymerases, Pol IV and Pol V. These plant-specific polymerases are not essential for viability but have nonredundant roles in RNA-mediated gene-silencing pathways. Proteomic analyses have revealed that *Arabidopsis* Pol IV and Pol V have a 12-subunit composition like Pol II. In fact, half of the subunits of Pols II, IV, and V are encoded by the same genes. The remaining Pol IV– or Pol V–specific subunit genes arose through duplication and subfunctionalization of ancestral Pol II subunit genes. These include the genes encoding the largest subunits unique to Pol IV or Pol V, the genes encoding the second- and the fourth-largest subunits that are used by both Pol IV and Pol V, the gene encoding the fifth-largest subunit unique to Pol V and the genes encoding the seventh-largest subunits that are unique to Pol IV and Pol V. On the basis of phylogenetic reconstructions, the gene duplication events giving rise to the first-, second-, fourth-, fifth-, and seventh-largest subunits of Pol IV and/or Pol V occurred independently. Interestingly, a cDNA-mediated duplication of the Pol II seventh-largest subunit gene via retrotranposition was an early event in Pol IV evolution, preceded only by the duplications of the largest and second-largest subunit genes. Secondary duplication of this cDNA-like gene to generate Pol IV– and Pol V–specific seventh-largest subunits has occurred in *Arabidopsis* but not all dicotyledonous plants or monocots, indicative of the dynamic evolution of RNA Pol IV and Pol V in plants.

Eukaryotes have three essential multisubunit RNA polymerases with unique and evolutionarily conserved functions, namely, RNA polymerases I, II, and III (Pols I, II, and III) (Sentenac 1985; Cramer et al. 2008). These complex enzymes, composed of 12–17 subunits, are found in all known metazoans, plants, and protists studied to date. Pol I transcribes 45S rRNA genes (Paule and White 2000; Russell and Zomerdijk 2006), Pol II transcribes protein-coding genes and a variety of noncoding RNAs (Hahn 2004), and Pol III mostly transcribes tRNA and 5S rRNA genes (Paule and White 2000; Schramm and Hernandez 2002). As first shown in yeast, Pols I, II, and III have a common set of subunits, encoded by the *RPB5*, *RPB6*, *RPB8*, *RPB10*, and *RPB12* genes (Woychik et al. 1990), as well unique subunits that include the largest and second-largest subunits, whose interaction forms the catalytic center for DNA-templated RNA polymerization (Cramer et al. 2008).

Sequencing of the *Arabidopsis thaliana* genome revealed the first evidence that plants have multisubunit RNA polymerases in addition to Pols I, II, and III. Specifically, genes for the largest and second-largest subunits of what turned out to be Pol IV (formerly Pol IVa) (Herr et al. 2005; Onodera et al. 2005) and Pol V (formerly Pol IVb) (Kanno et al. 2005; Pontier et al. 2005), were identified during the genome annotation process by Pikaard and Eisen (*Arabidopsis* Genome Initiative 2000; Pikaard et al. 2008). Orthologous genes were subsequently identified in the rice genome (Goff et al. 2002). Pol IV and Pol V are not essential for viability but have important roles in RNA-mediated silencing of repetitive elements, transposons, and transgenes, including cell-to-cell and long-distance spread or perception of silencing signals (Baulcombe 2006; Brodersen and Voinnet 2006; Pikaard et al. 2008; Matzke et al. 2009; Lahmy et al. 2010). Pol IV and Pol V have important roles in siRNA-directed DNA methylation and heterochromatin formation (Herr et al. 2005; Onodera et al. 2005; Lahmy et al. 2010). In the siRNA-directed DNA methylation pathway, Pol IV is necessary for the biogenesis of small RNAs required for gene silencing (Herr et al. 2005; Onodera et al. 2005; Zhang et al. 2007; Mosher et al. 2008). Pol V is not required for siRNA biogenesis at most loci (Mosher et al. 2008) but generates noncoding transcripts at loci where siRNA-directed DNA methylation occurs (Wierzbicki et al. 2008, 2009). siRNA-ARGONAUTE 4 complexes can be cross-linked to Pol V transcripts, indicating that Pol V transcripts serve as scaffolds for recruiting silencing machinery to the adjacent chromatin (Wierzbicki et al. 2009). The loci at which Pol IV or Pol V participate in silencing mostly overlap, but each polymerase has been implicated

[3]*Present address:* Department of Biochemistry, University of Wisconsin, Madison, WI 53706.

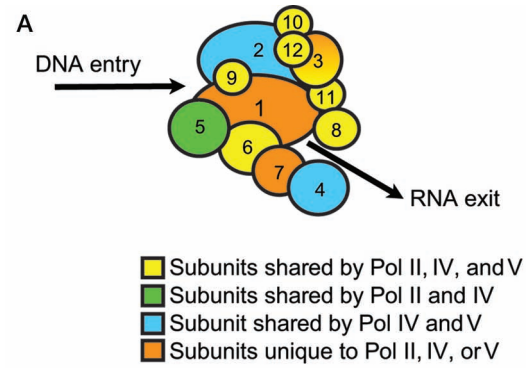
Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

in silencing at sites where the other is apparently not necessary (Zhang et al. 2007; Mosher et al. 2008; Douet et al. 2009; Pontes et al. 2009). Although Pol IV and Pol V have been studied most intensively in *Arabidopsis*, recent work in maize has revealed a prominent role for Pol IV and the RNA-directed DNA methylation pathway in the epigenetic phenomenon paramutation, as well as various aspects of maize development (Sidorenko and Chandler 2008; Erhard et al. 2009; Sidorenko et al. 2009; Hollick 2010) .

RNA polymerase subunits are typically encoded by a single gene in yeast and animals, whereas small multigene families exist for most plant nuclear RNA polymerase subunits (Ream et al. 2009). Phylogenetic analyses of the genes encoding the largest subunits of Pol IV and Pol V indicated that they likely arose stepwise, beginning with the duplication of the Pol II largest subunit gene (*NRPB1*) hundreds of millions of years ago in a common ancestor of land plants and the green alga Charales (Luo and Hall 2007). This initial event gave rise to the largest subunit of Pol IV (NRPD1). Subsequent duplication of the Pol IV largest subunit gene then gave rise to the Pol V largest subunit gene. Pol IV– or Pol V–specific second-largest subunit genes are not apparent in algae but are present in all land plants examined to date (Luo and Hall 2007). In *Arabidopsis*, the second-largest subunits of Pol IV and Pol V are encoded by a single gene (*NRPD2/ NRPE2*) and are more similar in sequence to the Pol II second-largest subunit (NRPB2) than to the paralogous subunits of Pol I or Pol III (Onodera et al. 2005; Luo and Hall 2007). These observations suggest that the earliest form of Pol IV probably involved a unique largest subunit paired with second-largest and noncatalytic subunits of Pol II (Luo and Hall 2007). Shared ancestry of Pols II, IV, and V is consistent with the fact that Pol II is required for small RNA-mediated heterochromatin formation and silencing in fission yeast (Buhler and Moazed 2007; Grewal and Elgin 2007; Zaratiegui et al. 2007). In plants, Pol IV and Pol V appear to have taken over most of these silencing functions, although recent studies have shown that Pol II continues to have a role in RNA-mediated silencing in plants (Zheng et al. 2009).

Using tandem mass spectrometry, we recently determined the compositions of affinity-purified *Arabidopsis* Pols II, IV, and V, showing that each has 12 core subunits (Ream et al. 2009). Multiple subunits of *Brassica oleracea* Pol V have also been identified (Huang et al. 2009) and are consistent with the *Arabidopsis* data. Approximately half of the subunits of Pol II, Pol IV, and Pol V are encoded by the same genes in *Arabidopsis* (Fig. 1) (Ream et al. 2009). The remaining distinct subunits of Pol IV and Pol V are clearly paralogs of Pol II subunits, confirming the common evolutionary history of Pols II, IV, and V (Ream et al. 2009).

Our recent determination of the Pol IV and Pol V subunit compositions prompted us to ask whether those subunits likely to account for the unique properties of Pol IV and Pol V arose in multiple steps over time, or possibly originated via concerted duplication events that facilitated a rapid fixation of Pol IV and Pol V functions. To address the question, we performed phylogenetic analyses of subunit families encoding the fourth-, fifth-, and seventh-largest subunits of polymerases II, IV, or V. The fifth-largest subunit of Pol II is implicated in interactions with transcription factors and the DNA template downstream from the polymerase active site (Lin et al. 1997; Gnatt et al. 2001; Wei et al. 2001; Cramer et al. 2008). The fourth- and seventh-largest subunits form a subcomplex that protrudes as a stalk from the polymerase core, interacting with the RNA as it exits the polymerase and mediating a variety of cotranscriptional and processing events (Pillai et al. 2001; Choder 2004; Ujvari and Luse 2006; Jasiak et al. 2008; Runner et al. 2008; Chen et al. 2009).



B

Subunit	Pol II	Pol IV	Pol V
Four	NRPB4	NRPD4/NRPE4	
Five	NRPB5/NRPD5		NRPE5
Seven	NRPB7	NRPD7	NRPE7

Figure 1. Identities and differences among Pols II, IV, and V subunits in *A. thaliana*. (*A*) The 12 core subunits are numbered according to the convention for yeast Pol II, with the largest subunit numbered 1. Differing subunit coloration reflects subunit usage in RNA Pols II, IV, and/or V, as indicated. (*B*) Nomenclature for the fourth-, fifth-, and seventh-largest subunits of Pols II, IV, and V in *Arabidopsis*.

Our analyses indicate that the genes encoding the fifth-, fourth-, and seventh-largest subunits of Pol IV and Pol V arose independently during the course of plant evolution and arose by distinct mechanisms. The results suggest that the functional divergence of Pols II, IV, and V has occurred through the additive effects of independent gene duplication events and subunit subfunctionalization processes likely to affect template selection, transcriptional regulation, and/or interactions with RNA.

METHODS

Phylogenetic Analyses

We determined the phylogenetic relationships among seventh-, fourth-, and fifth-largest subunits of Pols II, IV, and V by BLAST, searching against the "green plants"

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

portion of the NCBI ref_seq protein database. Only proteins with a similarity score of $e = 0.01$ or better and 60% or better coverage were subjected to further analyses. Proteins with multiple entries and proteins more than twice as long as the query sequence were removed manually. An additional BLAST search was performed with each query against the entire NCBI ref_seq protein database to obtain nonplant sequences that could be added to the analyses. Amino acid sequences were multiply aligned using MUSCLE version 3.7 with the default parameters (Edgar 2004). ProtTest version 2.4 was used to determine the best model of evolution for the resulting alignment by the Akaike Information Criterion (AIC) and maximum likelihood (Drummond and Strimmer 2001; Guindon and Gascuel 2003; Abascal et al. 2005). Phylogenetic trees were then inferred by maximum-likelihood (ML) analysis using RAxML version 7.0.4 (Stamatakis 2006; Stamatakis et al. 2008) on the CIPRES web portal (http://www.phylo.org/sub_sections/portal). Trees were generated independently three times, with 1000 bootstrap replicates in each run. Bipartition files were input to Geneious 5.0 (http://www.geneious.com/) to visualize the results of each of the three runs. Additional trees were generated using MrBayes version 3.1 at the CIPRES web portal with default settings to compare to ML results (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003; Miller et al. 2009).

Seventh-Largest Subunit Parameters

Pol II seventh-largest subunit BLAST searches (176 amino acid queries) yielded results for known RNA Pol II, Pol III, Pol IV, and Pol V subunit proteins. Nonplant sequences for Pol II and Pol III subunits were used in downstream analyses to further resolve the relationships of unknown plant subunit sequences. A total of 45 protein sequences were identified from algae, moss, monocotyledonous, and dicotyledonous flowering plants, as well as nonplants. Analyses in ProtTest 2.4 (http://darwin.uvigo.es/software/prottest.html) determined the best evolutionary model for this data set to be WAG.

Fourth-Largest Subunit Parameters

BLAST searches for fourth-largest subunits were initially performed using the 138-amino-acid NRPB4 (Pol II) sequence against the "green plant" entries in the NCBI ref_seq protein database. However, very few hits were obtained using these parameters, and *A. thaliana* NRPD4/NRPE4 (Pol IV/Pol V) was not among them, possibly because of the use of a short query sequence. To add Pol IV/Pol V fourth-largest subunit proteins to the data set, a second BLAST search was conducted using the NRPD4/NRPE4 amino acid sequence (Table 1 on following pages). NPRD4/NRPE4 has two predicted amino acid sequences, listed as AAT71989 and NP_193330 in the NCBI protein database. NRPD4/NRPE4 is annotated in TAIR9 as At4g15950, which corresponds to the protein sequence NP_193330. PCR-based cloning and sequencing of the NRPD4/NRPE4 cDNA from *A. thaliana* showed that the AAT71989 sequence is correct and exists in planta (data not shown), thus this longer sequence was used for subsequent phylogenetic analyses. Nonplant NRPB4 sequences were also included in the study. Amino acid sequences with lengths more than twice those of *Arabidopsis* NRPD4/NRPE4 were removed from the resulting combined data set. Analyses in ProtTest 2.4 determined the best evolutionary model for this data set to be JTT (Abascal et al. 2005).

Fifth-Largest Subunit Parameters

For fifth-largest subunits, the BLAST search used a 228-amino-acid NRPB5 (Pol II) sequence. ProtTest indicated that the appropriate evolutionary model to use for ML analysis with this data set was WAG.

RESULTS

Small Multigene Family Encodes Seventh-Largest Subunits of RNA Pols II, IV, and/or V

Proteomic analyses have identified unique seventh-largest subunits in RNA Pols II, IV, and V of *A. thaliana* (Ream et al. 2009), potentially contributing to the unique functions of the three enzymes. Using a Pol II seventh-largest subunit as a query sequence to find homologous proteins in GenBank (Table 1), followed by multiple alignments, maximum-likelihood phylogenetic trees reveal three classes of proteins (Fig. 2A). These three classes correspond to RNA Pol II seventh-largest subunit sequences of plants, fungi, and animals; RNA Pol III subunit sequences of plants, fungi, and animals; and a plant-specific class that includes the seventh-largest subunits of *A. thaliana* Pol IV or Pol V. The Pol II seventh-largest subunit class contains one protein sequence from each plant species represented in our data set. Branch lengths are generally shorter for Pol II seventh-largest subunit sequences than for other proteins in the tree, suggesting that this subunit sequence is more highly constrained than corresponding Pol IV or Pol V subunits, reflecting the highly conserved and essential functions of Pol II in eukaryotes.

A. thaliana Pol IV and Pol V seventh-largest subunits group within a class distinct from the Pol II and Pol III subunits, with high bootstrap support. This Pol IV/V class includes one sequence from the moss *Physcomitrella patens*, but no algal sequences, suggesting that the ancestral gene for Pol IV/Pol V seventh-largest subunits arose after the divergence of moss and algae but before the radiation of vascular plants. Outer branches of the tree indicate that further duplication events took place in some dicots, giving rise to distinct Pol IV– and Pol V–associated seventh-largest subunits in only some species, such as *Arabidopsis*. In *A. thaliana*, these Pol IV– and Pol V–specific genes are *NRPD7* and *NRPE7*, respectively (Ream et al. 2009).

A third *A. thaliana* protein of unknown function groups with the Pol IV and Pol V seventh-largest subunits. This unknown protein is encoded by *A. thaliana* gene *At4g14520* but has not been detected thus far in our analy-

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Table 1. Protein sequences used in the phylogenetic analyses of this study

Seventh-largest subunits

Sequence ID on tree	NCBI accession number	Species
A. aegypti	XP_001661998	*Aedes aegypti*
A. mellifera	XP_394665	*Apis mellifera*
A. thaliana At1g06790	NP_973776	*Arabidopsis thaliana*
A. thaliana NRPB7	NP_200726	*Arabidopsis thaliana*
A. thaliana NRPE7	NP_193202	*Arabidopsis thaliana*
A. thaliana NRPD7	NP_566719	*Arabidopsis thaliana*
A. thaliana At2g14520	NP_193188	*Arabidopsis thaliana*
A. thaliana RPC25	NP_172164	*Arabidopsis thaliana*
C. elegans	NP_505625	*Caenorhabditis elegans*
C. reinhardtii 1	XP_001692125	*Chlamydomonas reinhardtii*
C. reinhardtii 2	XP_001692857	*Chlamydomonas reinhardtii*
G. max	P46279	*Glycine max*
H. sapiens	NP_612211	*Homo sapiens*
M. mulatta	XP_001105258	*Macaca mulatta*
M. truncatula 1	ACJ84197	*Medicago truncatula*
M. truncatula 2	ACJ85752	*Medicago truncatula*
M. pusilla	EEH57697	*Micromonas pusilla CCMP1545*
M. sp. RCC299 1	ACO61449	*Micromonas sp. RCC299*
M. sp. RCC299 2	ACO61588	*Micromonas sp. RCC299*
O. sativa J 1	NP_001055483	*Oryza sativa japonica* group
O. sativa J 2	NP_001059628	*Oryza sativa japonica* group
O. sativa J 3	NP_001054703	*Oryza sativa japonica* group
O. lucimarinus	XP_001420140	*Ostreococcus lucimarinus CCE9901*
O. tauri	CAL53271	*Ostreococcus tauri*
P. patens 1	XP_001780058	*Physcomitrella patens* subsp. *patens*
P. patens 2	XP_001756041	*Physcomitrella patens* subsp. *patens*
P. patens 3	XP_001751839	*Physcomitrella patens* subsp. *patens*
P. trichocarpa 1	XP_002298912	*Populus trichocarpa*
P. trichocarpa 2	XP_002300802	*Populus trichocarpa*
P. trichocarpa 3	XP_002312568	*Populus trichocarpa*
P. trichocarpa 4	ABK95475	*Populus trichocarpa*
P. trichocarpa 5	XP_002314653	*Populus trichocarpa*
R. communis 1	EEF41911	*Ricinus communis*
R. communis 2	EEF39694	*Ricinus communis*
S. cerevisiae 1	NP_010692	*Saccharomyces cerevisiae*
S. cerevisiae 2	NP_012778	*Saccharomyces cerevisiae*
S. aucuparia	ABH06364	*Sorbus aucuparia*
V. vinifera 1	XP_002282298	*Vitis vinifera*
V. vinifera 2	XP_002281971	*Vitis vinifera*
V. vinifera 3	XP_002271507	*Vitis vinifera*
V. vinifera 4	XP_002275475	*Vitis vinifera*
V. vinifera 5	XP_002284221	*Vitis vinifera*
Z. mays 1	NP_001149597	*Zea mays*
Z. mays 2	NP_001147612	*Zea mays*
Z. mays 3	NP_001150375	*Zea mays*

Fourth-largest subunits

Sequence ID on tree	NCBI accession number	Species
A. arenosa	ACK44499	*Arabidopsis arenosa*
A. thaliana NRPD4/E4	AAT71989	*Arabidopsis thaliana*
A. thaliana NRPB4	NP_196554	*Arabidopsis thaliana*
A. annua	ABQ32301	*Artemisia annua*
B. taurus	NP_001069926	*Bos taurus*
B. rapa	AAZ66921	*Brassica rapa*
C. elegans	NP_495544	*Caenorhabditis elegans*
D. rerio	NP_001002317	*Danio rerio*
D. discoideum	XP_640235	*Dictyostelium discoideum* AX4
D. melanogaster	NP_001014633	*Drosophila melanogaster*
E. guineensis	ACF06509	*Elaeis guineensis*
H. sapiens	NP_004796	*Homo sapiens*
M. truncatula	ABN08874	*Medicago truncatula*
M. pusilla	EEH58834	*Micromonas pusilla* CCMP1545
M. sp. RCC299	ACO67222	*Micromonas* sp. RCC299

(*Continued on facing page.*)

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Table 1. (*Continued*)

Sequence ID on tree	NCBI accession number	Species
M. musculus	NP_081377	*Mus musculus*
O. sativa J 1	NP_001045683	*Oryza sativa japonica* group
O. sativa J 2	NP_001047971	*Oryza sativa japonica* group
O. lucimarinus	XP_001419467	*Ostreococcus lucimarinus* CCE9901
P. patens	XP_001766952	*Physcomitrella patens* subsp. *patens*
P. trichocarpa 1	XP_002307132	*Populus trichocarpa*
P. trichocarpa 2	XP_002329936	*Populus trichocarpa*
R. communis 1	EEF37360	*Ricinus communis*
R. communis 2	EEF40418	*Ricinus communis*
S. cerevisiae	NP_012395	*Saccharomyces cerevisiae*
V. vinifera 1	CAO40474	*Vitis vinifera*
V. vinifera 2	XP_002282413	*Vitis vinifera*
X. laevis	NP_001087315	*Xenopus laevis*
Z. mays 1	NP_001130236	*Zea mays*
Z. mays 2	NP_001148722	*Zea mays*

Fifth-largest subunits

Sequence ID on tree	NCBI accession number	Species
A. thaliana NRPE5	NP_191267	*Arabidopsis thaliana*
A. thaliana 1	BAB02760	*Arabidopsis thaliana*
A. thaliana NRPB5/ NRPD5	NP_188871	*Arabidopsis thaliana*
A. thaliana At2g413140	NP_181665	*Arabidopsis thaliana*
A. thaliana At5g57980	NP_200606	*Arabidopsis thaliana*
A. thaliana At3g54490	NP_191013	*Arabidopsis thaliana*
B. napus	AAF81222	*Brassica napus*
C. elegans	NP_491961	*Caenorhabditis elegans*
C. reinhardtii	XP_001697601	*Chlamydomonas reinhardtii*
D. rerio	NP_001003564	*Danio rerio*
D. discoideum	XP_635268	*Dictyostelium discoideum* AX4
D. melanogaster	NP_610630	*Drosophila melanogaster*
E. guineensis	ACF06443	*Elaeis guineensis*
H. sapiens	NP_002686	*Homo sapiens*
M. truncatula 1	ABN07995	*Medicago truncatula*
M. truncatula 2	ABD28306	*Medicago truncatula*
M. pusilla	EEH55227	*Micromonas pusilla* CCMP1545
M. sp RCC299	ACO66413	*Micromonas sp.* RCC299
M. musculus	NP_079830	*Mus musculus*
O. sativa J 1	NP_001044564	*Oryza sativa japonica* group
O. sativa J 2	NP_001065723	*Oryza sativa japonica* group
O. sativa J 3	NP_001066119	*Oryza sativa japonica* group
O. sativa J 4	NP_001053145	*Oryza sativa japonica* group
O. lucimarinus	XP_001417617	*Ostreococcus lucimarinus* CCE9901
P. patens 1	XP_001758465	*Physcomitrella patens* subsp. *patens*
P. patens 2	XP_001754558	*Physcomitrella patens* subsp. *patens*
P. patens 3	XP_001770728	*Physcomitrella patens* subsp. *patens*
P. sitchensis 1	ABK23892	*Picea sitchensis*
P. sitchensis 2	ABR18148	*Picea sitchensis*
P. abelii	NP_001126823	*Pongo abelii*
P. trichocarpa 1	XP_002323257	*Populus trichocarpa*
P. trichocarpa 2	XP_002297750	*Populus trichocarpa*
P. trichocarpa 3	XP_002331171	*Populus trichocarpa*
R. communis 1	EEF48219	*Ricinus communis*
R. communis 2	EEF49580	*Ricinus communis*
R. communis 3	EEF33224	*Ricinus communis*
S. cerevisiae	NP_009712	*Saccharomyces cerevisiae*
V. vinifera 1	XP_002269723	*Vitis vinifera*
V. vinifera 2	XP_002275071	*Vitis vinifera*
V. vinifera 3	XP_002284107	*Vitis vinifera*
X. tropicalis	NP_001016473	*Xenopus* (*Silurana*) *tropicalis*
Z. mays 1	NP_001141164	*Zea mays*
Z. mays 2	NP_001141853	*Zea mays*
Z. mays 3	NP_001132429	*Zea mays*

The table provides National Center for Biotechnology Information (NCBI) accession numbers for protein sequences used for the generation of phylogenetic trees. The name for each protein within the trees is provided in the first column.

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

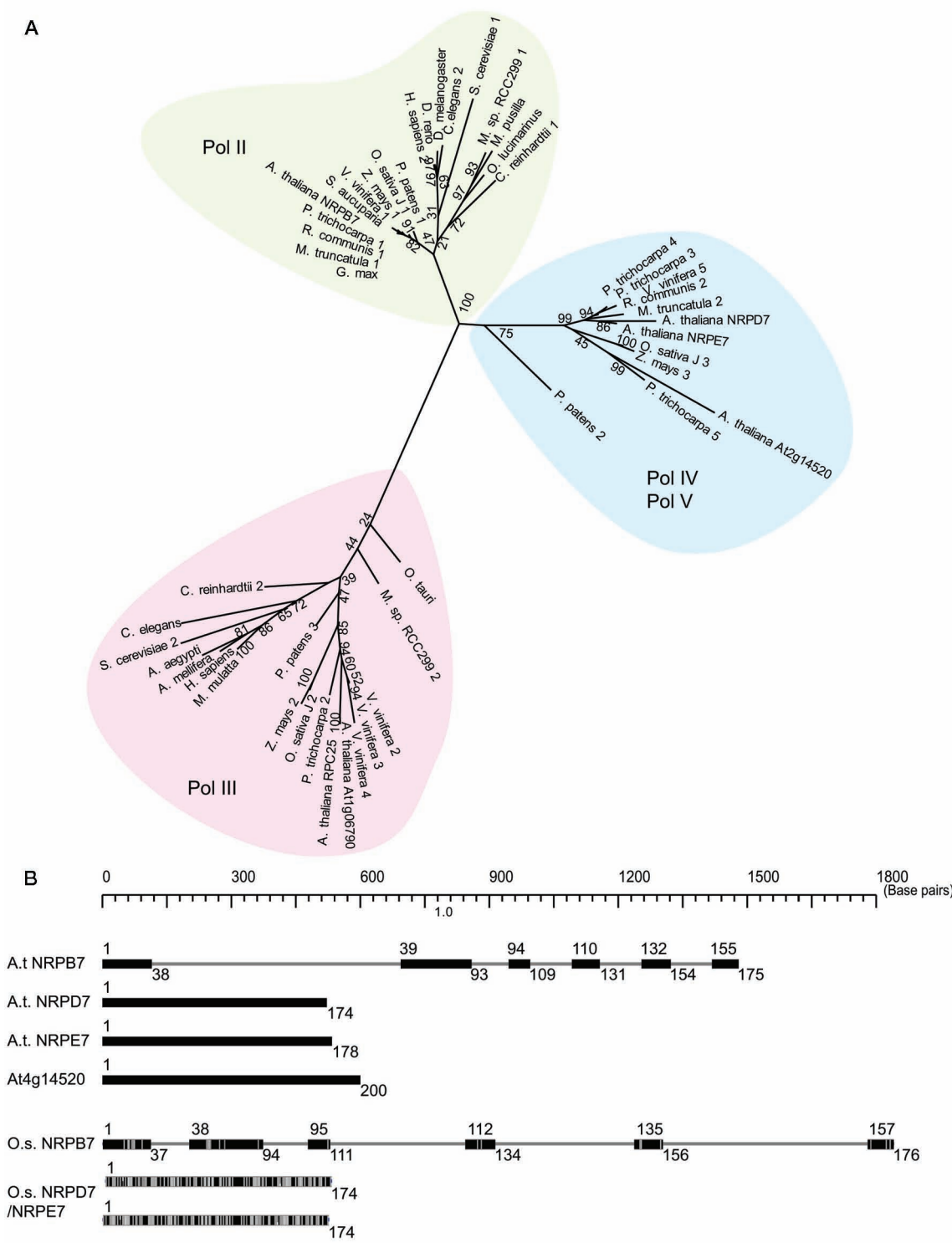


Figure 2. Pol IV and V seventh-largest subunits are encoded by intronless genes likely to have arisen through retrotransposition in a common ancestor of moss and higher plants. (*A*) Maximum-likelihood phylogenetic tree inferred from multiple alignments of amino acid sequences. Proteins similar to seventh-largest subunits of Pol II are aligned, including plant, budding yeast, *Drosophila*, *Caenorhabditis elegans*, and human sequences. Numbers at nodes in the tree are bootstrap values from 1000 replicates. (*B*) Gene structures for seventh-largest subunits of Pol II (NRPB7), Pol IV (NRPD7), and Pol V (NRPE7) in *A. thaliana* and rice (*Oryza sativa japonica)*. (Horizontal lines) Introns, (filled boxes) exons. Numbers above and below boxes indicate approximate amino acid positions at beginnings and ends of exons, respectively. Gene models shown for rice were modified from TARGeT outputs. Increasingly dark bars indicate increasing similarity with *A. thaliana* query sequences used in analyses (*A. thaliana* NRPB7 and NRPD7).

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

ses of Pols II, IV, or V by mass spectrometry (Ream et al. 2009). However, it is intriguing that an ortholog of this unknown protein is also present in poplar (*Populus trichocarpa*), with high bootstrap support. The placement of *At4g14520* and its poplar homolog is not well resolved within the Pol IV/Pol V seventh-largest subunit group, and no sequences grouping with these two proteins were identified in other plants. The relatively long branches associated with these two sequences may have led to their erroneous inclusion in the subgroup containing monocot sequences. Whether the protein encoded by *At4g14520* associates with other polymerase subunits is not known.

Pols IV and V Seventh-Largest Subunit Genes Arose through Retrotransposition of Pol II Seventh-Largest Subunit mRNA

Comparison of the *A. thaliana NRPB7* (Pol II), *NRPD7* (Pol IV), *NRPE7* (Pol V), and *At4g14520* (unknown) gene models reveals that introns are absent from the latter three genes, whereas five introns are present in *NRPB7* (Fig. 2B). Gene structure comparisons across additional plant species were done using TARGeT (Han et al. 2009), confirming the lack of introns in genes corresponding to the Pol IV/Pol V seventh-largest subunit group. The absence of introns in Pol IV/Pol V seventh-largest subunit coding regions, in both monocots and dicots, suggests that their common ancestral gene was generated by copying the Pol II seventh-largest gene (*NRPB7*) mRNA into cDNA via the action of RNA-dependent DNA polymerase (reverse transcriptase), followed by incorporation of the cDNA into the genome and acquisition of promoter and other regulatory sequences.

Arabidopsis Pol IV– and V–Specific Seventh-Largest Subunits Diverged Recently following Segmental Duplication

The duplication event allowing for the divergence of Pol IV– and Pol V–specific seventh-largest subunit proteins from their common ancestor appears to have occurred very recently in *Arabidopsis*, as the genes coding for NRPD7 (Pol IV) and NRPE7 (Pol V) seventh-largest subunit proteins are sister to each other within the larger dicot Pol IV/Pol V subgroup (Fig. 2A). A similar duplication event took place independently in poplar after the divergence of *Arabidopsis* and poplar from each other and from other sampled dicots. A third *A. thaliana* polypeptide sequence encoded by *At2g14520* groups with a poplar sequence within the Pol IV/Pol V class. The polymerase association of this protein, if any, is not known, as this protein has not been identified by mass spectrometry to date. Only one protein encoding a Pol IV/Pol V seventh-largest subunit is detected in *Vitis vinifera*, *Ricinus communis*, or *Medicago truncatula*. The lack of a second protein sequence from these species suggests a lack of Pol IV– and Pol V–specific seventh-largest subunits in the proteomes of these plants or may simply be the result of incomplete protein sequence information available for these species.

The recent divergence of Pol IV and Pol V seventh-largest subunit proteins was further investigated using the TIGR segmental duplication tool (ftp://ftp.tigr.org/pub/software/Blast-Synteny-Toolkit/ArabDups_n_XYplotter.tar.gz), identifying the Pol IV/Pol V seventh-largest subunit genes as segmentally duplicated genes. This duplication is likely to have occurred as part of a recent genome duplication, followed by chromosomal rearrangement. This polyploidy event is hypothesized to have occurred before the split between *Brassica* and *Arabidopsis* but after the split between *Arabidopsis* and cotton (Blanc et al. 2003). Therefore, any functional differences between Pol IV and Pol V imparted by having distinct, specialized seventh-largest subunits were probably acquired relatively recently and only in some dicots, and possibly only in *Arabidopsis* and its close relatives.

Functional Divergence of Fourth-Largest Subunit Genes

Two distinct fourth-largest subunit proteins were identified in proteomic studies of *Arabidopsis* RNA Pols II, IV, and V. Pol II uses NRPB4, whereas Pol IV and Pol V use a fourth-largest subunit protein, NRPD4/NRPE4, that is encoded by the same gene (He et al. 2009; Ream et al. 2009).

Maximum-likelihood phylogenetic trees show two major groupings of fourth-largest subunit proteins, one corresponding to known Pol II subunits and another containing the *Arabidopsis* Pol IV/Pol V subunit sequence (NCBI: AAT71989) (Fig. 3). The class of sequences containing *A. thaliana* NRPD4/NRPE4 (Pol IV/Pol V) splits with high boot strap support from the poorly resolved Pol II fourth-largest subunit sequences and includes proteins from both monocots and dicots, but not moss, consistent with a duplication of the Pol II fourth-largest subunit gene after flowering plants split from moss, but before the divergence of monocots and dicots. The phylogenetic relationships among plant Pol II fourth-largest subunit sequences are not well resolved.

The *A. thaliana* and *Arabidopsis lyrata* Pol II fourth-largest subunit genes contain five introns, whereas their Pol IV/Pol V genes (*NRPD4/NRPE4*) have four (Fig. 4B). The Pol II and Pol IV/Pol V fourth-largest subunit genes were not found to be related by virtue of a whole-genome duplication, based on analyses using the TIGR segmental duplication tool. Instead, their origin traces to a segmental duplication.

Fifth-Largest Subunit Gene Duplicated Multiple Times, Leading to Further Functional Differentiation of Pol V from Pols II and IV

Maximum-likelihood trees show that fifth-largest subunit proteins group into two major classes, with strong bootstrap support (Fig. 4A). One class contains Pol II subunits of diverse species including the *A. thaliana* gene encoding the subunit used by Pols I, II, III, and IV and therefore having the synonymous names NRPA5, NRPB5,

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

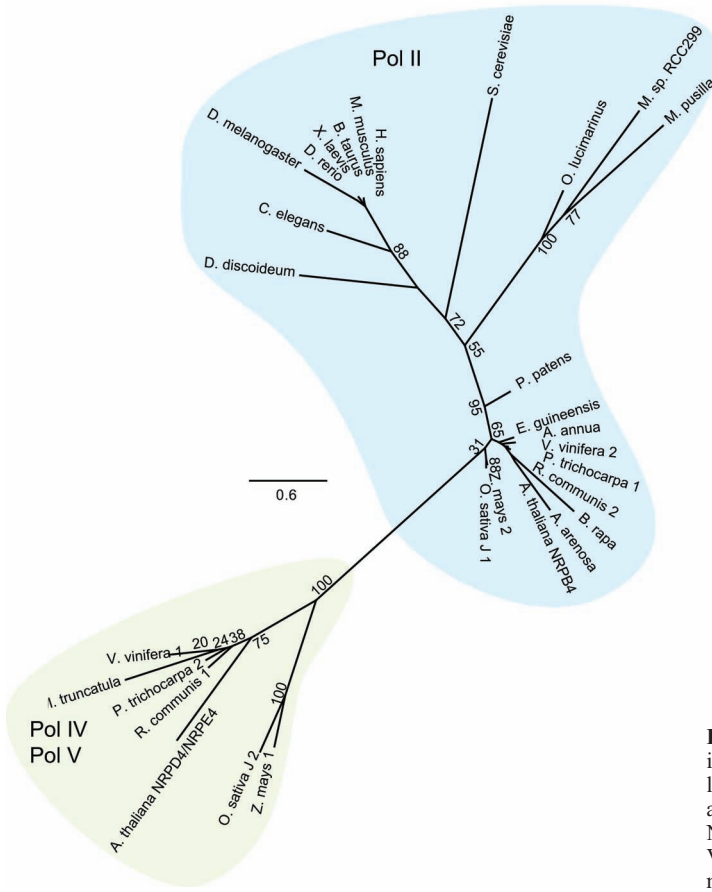


Figure 3. Maximum-likelihood phylogenetic tree showing relationships among RNA Pols II, IV, and V fourth-largest subunits. Tree was created based on multiple alignment of protein sequences similar to *A. thaliana* NRPB4. Only plant sequences group within the Pol V/Pol V clade. Numbers are bootstrap values determined from 1000 replicates.

NRPC5, or NRPD5. Two additional *A. thaliana* proteins group with this Pol I/II/III/IV class. Both of these protein sequences, one encoded the by *A. thaliana* gene *At5g57980* and the other a pseudogene according to its GenBank entry (BAB02760), are on long branches relative to the rest of the tree, indicating the possibility of greater divergence than is observed for the highly constrained Pol I/II/III/IV subunit.

The second major class in the tree of fifth-largest subunits contains the *A. thaliana* Pol V subunit gene *NRPE5*. The Pol V class is separated from other fifth-largest subunits by a longer branch than is seen elsewhere in the tree, suggesting more amino acid changes between the two classes relative to changes within each functional class. The NRPE5 class includes genes of monocot and dicot angiosperms, but not moss or other lower plants. These data suggest a duplication event in angiosperms before the diversification of monocots and dicots. Both monocot sequences in the Pol V class group together with high bootstrap support. Two sequences were identified in *O. sativa* (rice), suggesting a recent duplication event, apparently after the divergence of rice from maize because maize has only one Pol V–like fifth-largest subunit.

Two additional duplication events took place in the Pol V fifth-largest subunit cluster. Both of these duplications occurred after the split between monocots and dicots, resulting in subclasses represented by NRPE5 and the protein encoded by *A. thaliana* gene *At3g54490* and the *A. thaliana* gene *At2g31430*, which have no known polymerase affiliation (Ream et al. 2009). A query of the TIGR *Arabidopsis* segmental duplication database indicates that the *NRPE5* (Pol V) gene and the *At2g31430* sister gene diverged as part of a recent whole-genome duplication, followed by chromosomal rearrangements (Simillion et al. 2002; Blanc et al. 2003; Ermolaeva et al. 2003). No species other than *A. thaliana* contains three protein sequences in the Pol V class of fifth-largest subunits.

Fifth-Largest Subunit Gene Structure Conserved across All Sampled Plants except *Medicago truncatula*

Comparison of *Arabidopsis* fifth-largest subunit gene structures suggests that the *NRPE5* (Pol V) class of genes arose through a series of segmental duplications (Fig. 4B). The *NRPB5/NRPD5* and all three *A. thaliana* genes encoding proteins in the Pol V class contain three introns. Interestingly, the *At5g57980* gene in the Pol I/II/III/IV group has only one intron (data not shown). The location of this intron does not match any of the locations of those found in other family members, indicating that *At5g57980* might have arisen from a duplication event by retrotransposition followed by the gain of a new intron.

TARGeT analysis was conducted to compare gene structures for *NRPB5/NRPD5* (Pol II/Pol IV) and *NRPE5* (Pol V) across a greater sample of plant species (Han et

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

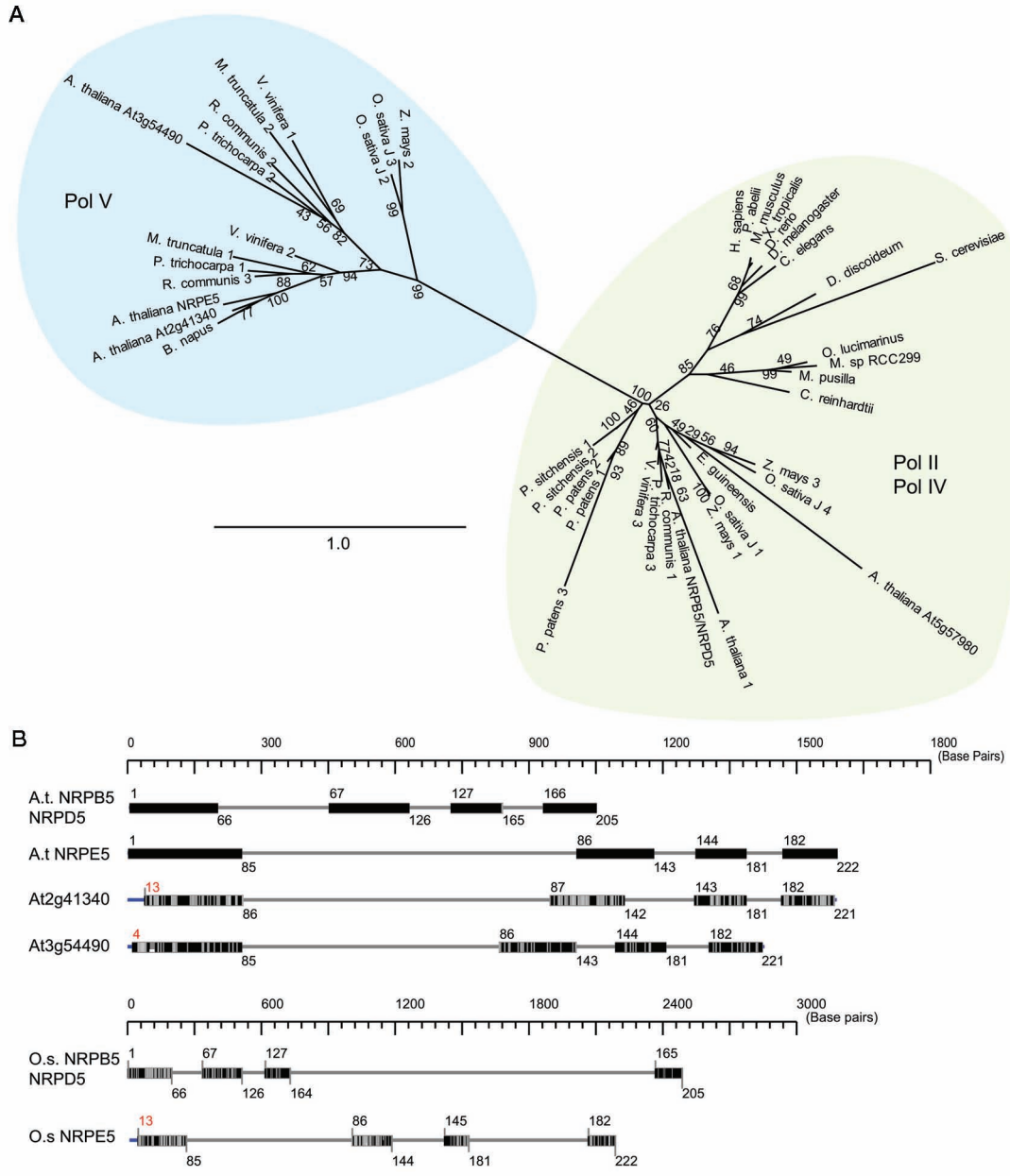


Figure 4. Relationships among Pols II, IV, and V fifth-largest subunits. (*A*) Maximum-likelihood phylogenetic tree based on fifth-largest subunit amino acid sequence alignments, revealing plant-specific Pol V (NRPE5) clade. Bootstrap values were calculated from 1000 replicates. (*B*) Comparison of gene structures for *A. thaliana* (A.t.) NRPB5/NRPD5, NRPE5, and two NRPE5-like proteins, as well as rice (O.s.) NRPB5 and NRPE5. Gene models were modified from TARGeT output files, with introns depicted by lines and exons by boxes. (Darker bars) Greater degrees of amino acid similarity to the query sequence (*A. thaliana* NRPB5/NRPD5 or NRPE5).

al. 2009). The analysis was performed using the amino acid sequences for both *A. thaliana* NRPB5/NRPD5 (Pol II/Pol IV) and NRPE5 (Pol V). *NRPB5/NRPD5* and *NRPE5* genes contain three introns in monocots and dicots, except for *M. truncatula,* in which all genes predicted by TARGeT have only two introns (data not shown). *A. lyrata* was the only species found to contain a single-intron gene similar in structure to *At5g57980* (data not shown), indicating that this gene arose via a duplication event in *Arabidopsis* or a close ancestor, probably as part of a recent polyploidization event (Blanc et al. 2003).

DISCUSSION

Determining the relationships among plant nuclear RNA polymerase subunit family members allows us to draw conclusions regarding the evolutionary history and functional diversification of Pol IV and Pol V. Our data support the hypothesis that multiple independent gene duplication events gave rise to families of RNA polymerase subunits that define Pols II, IV, and V (Fig. 5). Each subunit has a distinct role within the polymerase complex, thus the additive effect of having multiple unique subunits

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

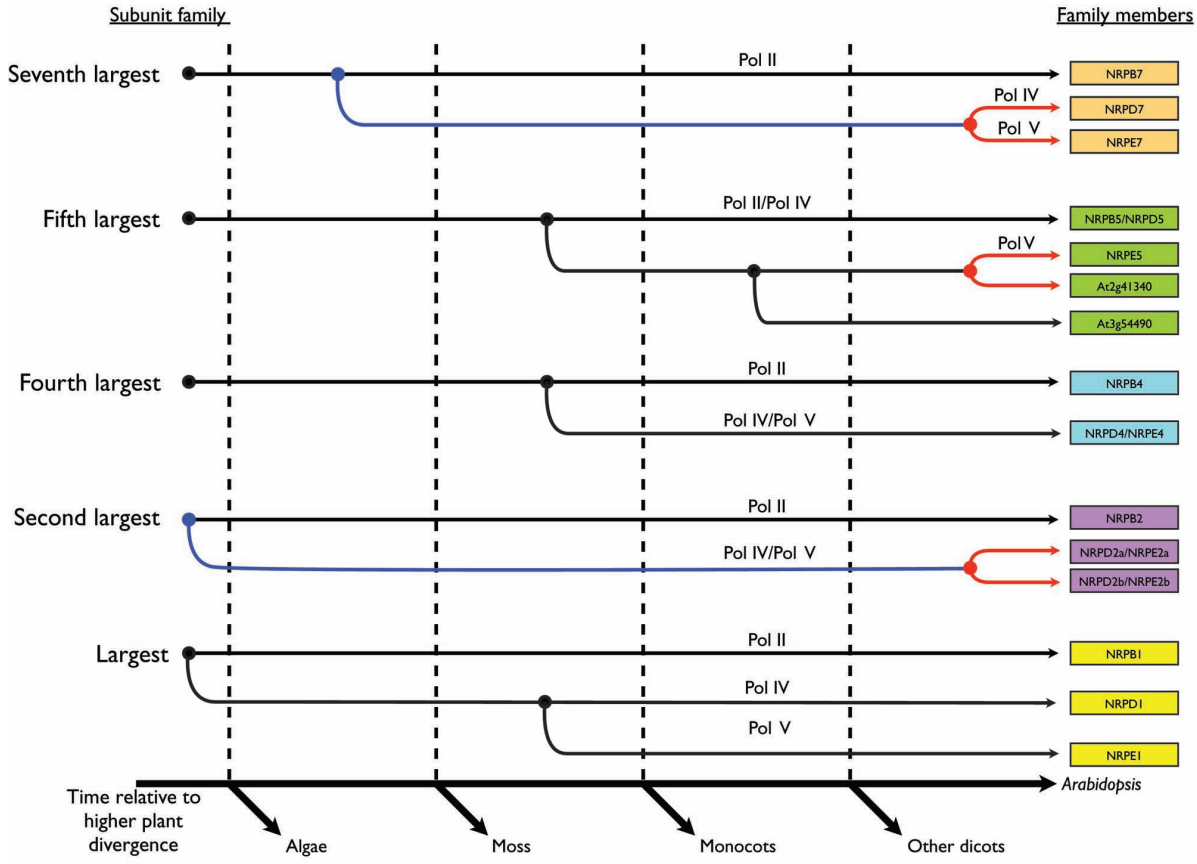


Figure 5. Gene duplication events giving rise to distinct subunits for RNA Pols II, IV, and/or V. Diagram shows relative timing of subunit duplications relative to major divergence events along a plant evolutionary time line. Filled circles and new branches indicate duplication events for genes still present in *A. thaliana*. (Red filled circles and lines) Duplications by segmental duplication, (blue circles and lines) duplications by retrotransposition. Subunit family members whose polymerase association is known are labeled above the lines representing their ancestral duplication events.

is likely to contribute to the functional diversification of the enzymes. Our analyses on the basis of the complete subunit compositions of Pol IV and Pol V extend and support the model put forth by Luo and Hall (2007), based on the analysis of only the two largest subunits, suggesting that Pols IV and V evolved via a multistep process. Because the protein subunits arose at different times, initial Pol IV/V subunits presumably formed functional complexes with subunits also used by Pol II, which remains the case for half of the Pol IV and Pol V subunits. It remains a possibility that Pol II subunits can substitute for noncatalytic Pol IV– and Pol V–specific subunits if the latter are lost because of mutation, a hypothesis that can be tested genetically and biochemically.

The fourth-largest subunit of RNA Pol II forms a subcomplex with the seventh-largest subunit, and this has been confirmed in plants (Larkin and Guilfoyle 1998). The yeast fourth-largest subunit has been shown to associate with transcription factors, presumably having a role in the sequence-specific recruitment of the transcription machinery to gene promoters (Orlicky et al. 2001). Seventh-largest subunit null mutations are lethal in yeast, but mutants bearing temperature-sensitive fourth-largest subunit alleles are viable and show defects in transcript elongation and 3′-end processing (Runner et al. 2008; Verma-Gaur et al. 2008). The unique seventh- and fourth-largest subunit subcomplexes of Pols II, IV, and V have interesting implications for the functional diversification of the plant-specific polymerases in terms of template selection and downstream RNA processing via different interactions with transcriptional activators, repressors, or RNA processing activities. NRPD7 (Pol IV) and NRPE7 (Pol V) appear to have diverged from each other relatively recently, after *A. thaliana* split from the other sampled dicots. The presence of a Pol IV/V-like seventh-largest subunit in moss suggests an important role for the seventh-largest subunit in the early functional diversification of Pols IV and V from Pol II.

The fourth-largest subunit of Pol IV and Pol V is encoded by the same gene in *A. thaliana* (He et al. 2009; Ream et al. 2009). Therefore, functional differences between Pol IV and Pol V are unlikely to be due to this subunit alone. However, the NRPB4–NRPB7, NRPD4/E4–NRPD7, and NRPD4/E4–NRPE7 subcomplexes are unique in all three complexes and could have a role in functional divergence of the plant RNA polymerases. Pro-

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

teins grouping with NRPD4/NRPE4 (Pols IV and V) are found in monocots and dicots but not mosses. It is possible that the moss sequence was missed in this analysis because of statistical limitations based on the short sequence length of the protein. However, less stringent searches also failed to identify a moss homolog. The lack of a Pol IV/V-specific fourth-largest subunit in moss suggests that Pols II, IV, and V use the product of the same gene. It cannot be ruled out that other species of moss might contain additional genes encoding polymerase subunits, but only one available bryophyte protein sequence database was readily available at the time of this analysis.

It is interesting that the fourth-largest subunit of Pol IV and Pol V appears to have arisen after the seventh-largest subunit, suggesting that the seventh-largest subunit may be more important in the diversification of Pol IV or Pol V functions relative to Pol II. The separate duplication and divergence of the fourth- and seventh-largest subunit families support the hypothesis that the functional divergences of Pol IV and Pol V from Pol II occurred in multiple steps with additive impacts.

The RNA Pol II fifth-largest subunit, NRPB5, interacts with and stabilizes the DNA template as it enters the polymerase. NRPB5 associates with the largest, catalytic subunit as well as the sixth-largest subunit, as revealed in yeast Pol II crystal structures (Cramer et al. 2008). Interactions between the Pol II fifth-largest subunit and several transcription factors have also been detected, implicating the subunit in target selection and transcriptional regulation (Miyao and Woychik 1998). A single fifth-largest subunit gene encodes the corresponding subunit of RNA Pols I, II, and III in yeast and humans (Woychik et al. 1990; Werner 2007). In *A. thaliana*, immunological and proteomics experiments have shown that the fifth-largest subunit of Pols I, II, III, and IV is encoded by the same gene with the synonymous names NRPA5/NRPB5/NRPC5/NRPD5 (Saez-Vasquez and Pikaard 1997; Larkin et al. 1999; Ream et al. 2009). However, Pol V has a unique fifth-largest subunit, NRPE5 (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). Therefore, whereas the fourth-largest subunits of Pol IV and Pol V are encoded by the same gene, distinct from the Pol II gene, the fifth-largest subunits of Pols II and IV are the same, but Pol V uses a distinct subunit. Several other genes have been identified as being similar to NRPB5/NRPD5 and NRPE5, but their direct association with particular polymerase complexes has not been reported.

The unique fifth-largest subunit of Pol V could be involved in selection of a unique template, possibly comprised of dsDNA, ssDNA, dsRNA, ssRNA, or some combination thereof. Phylogenetic analysis of the fifth-largest subunit family identified several predicted proteins grouping with NRPE5 (Pol V). However, no sequences grouping with NRPE5 (Pol V) were found in moss, indicating a radiation of the family in higher plants following the divergence of the common ancestor to the Pol IV and Pol V seventh-largest subunit proteins. The possible activities of these additional Pol V–like fifth-largest subunits are not known in higher plants, but their tissue-specific expression patterns indicate that the alternative fifth-largest subunits might interact with specific template molecules or activator proteins in different cell types (Ream et al. 2009).

The duplication events responsible for Pol IV– and Pol V–specific subunits appear to have taken place by segmental duplications and duplication by the reverse transcription of ancestral mRNA. Retrotransposition has also been proposed as the method of duplication for the second-largest subunit (Luo and Hall 2007). Several protein families underwent (additional) duplications in what appears to be a whole-genome duplication followed by chromosomal rearrangements in a recent ancestor of *Arabidopsis* (Fig. 5, red lines) (Blanc et al. 2003). The different modes of gene duplication at multiple points during plant evolution presumably allowed Pol IV and Pol V to derive unique functions over time. It will be interesting to use mass spectrometry to compare the subunit compositions of plant-specific nuclear RNA polymerases in monocots and moss, as well as to evaluate whether potentially different combinations of subunit variants are found in *Arabidopsis* Pol IV or Pol V at different times in development.

ACKNOWLEDGMENTS

We thank Stuart McDaniel and Ken Olsen for helpful conversations. This work was supported by National Institutes of Health grant GM57790 and the Monsanto–Washington University Plant Biology Collaborative Research Program.

REFERENCES

Abascal F, Zardoya R, Posada D. 2005. ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21:** 2104–2105.

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

Baulcombe DC. 2006. Short silencing RNA: The dark matter of genetics? *Cold Spring Harb Symp Quant Biol* **71:** 13–20.

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.

Brodersen P, Voinnet O. 2006. The diversity of RNA silencing pathways in plants. *Trends Genet* **22:** 268–280.

Buhler M, Moazed D. 2007. Transcription and RNAi in heterochromatic gene silencing. *Nat Struct Mol Biol* **14:** 1041–1048.

Chen CY, Chang CC, Yen CF, Chiu MT, Chang WH. 2009. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci* **106:** 127–132.

Choder M. 2004. Rpb4 and Rpb7: Subunits of RNA polymerase II and beyond. *Trends Biochem Sci* **29:** 674–681.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, et al. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* **37:** 337–352.

Douet J, Tutois S, Tourmente S. 2009. A Pol V-mediated silencing, independent of RNA-directed DNA methylation, applies to 5S rDNA. *PLoS Genet* **5:** e1000690. doi: 10.1371/journal.pgen.1000690.

Drummond A, Strimmer K. 2001. PAL: An object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17:** 662–663.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Erhard KF Jr, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB. 2009. RNA polymerase IV functions in paramutation in *Zea mays*. *Science* **323:** 1201–1205.

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Ermolaeva MD, Wu M, Eisen JA, Salzberg SL. 2003. The age of the *Arabidopsis thaliana* genome duplication. *Plant Mol Biol* **51:** 859–866.

Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292:** 1876–1882.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296:** 92–100.

Grewal SI, Elgin SC. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447:** 399–406.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Hahn S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11:** 394–403.

Han Y, Burnette JM III, Wessler SR. 2009. TARGeT: A web-based pipeline for retrieving and characterizing gene and transposable element families from genomic sequences. *Nucleic Acids Res* **37:** e78. doi: 10.1093/nar/gkp295.

He XJ, Hsu YF, Pontes O, Zhu J, Lu J, Bressan RA, Pikaard C, Wang CS, Zhu JK. 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev* **23:** 318–330.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308:** 118–120.

Hollick JB. 2010. Paramutation and development. *Annu Rev Cell Dev Biol* **26:** 557–579.

Huang L, Jones AM, Searle I, Patel K, Vogler H, Hubner NC, Baulcombe DC. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16:** 91–93.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Jasiak AJ, Hartmann H, Karakasili E, Kalocsay M, Flatley A, Kremmer E, Strasser K, Martin DE, Soding J, Cramer P. 2008. Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J Biol Chem* **283:** 26423–26427.

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37:** 761–765.

Lahmy S, Pontier D, Cavel E, Vega D, El-Shami M, Kanno T, Lagrange T. 2009. PolV (PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci* **106:** 941–946.

Lahmy S, Bies-Etheve N, Lagrange T. 2010. Plant-specific multisubunit RNA polymerase in gene silencing. *Epigenetics* **5:** 4–8.

Larkin RM, Guilfoyle TJ. 1998. Two small subunits in *Arabidopsis* RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J Biol Chem* **273:** 5631–5637.

Larkin RM, Hagen G, Guilfoyle TJ. 1999. *Arabidopsis thaliana* RNA polymerase II subunits related to yeast and human RPB5. *Gene* **231:** 41–47.

Lin Y, Nomura T, Cheong J, Dorjsuren D, Iida K, Murakami S. 1997. Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. *J Biol Chem* **272:** 7132–7139.

Luo J, Hall BD. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* **64:** 101–112.

Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ. 2009. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* **21:** 367–376.

Miller MA, Holder MT, Vos R, Midford PE, Liebowitz T, Chan L, Hoover P, Warnow T. 2009. The CIPRES portals. CIPRES. (http://www.phylo.org/sub_sections/ portal. Accessed: 2009-08-04, archived by WebCite® at http://www.webcitation.org/5imQ1JeQa.)

Miyao T, Woychik NA. 1998. RNA polymerase subunit RPB5 plays a role in transcriptional activation. *Proc Natl Acad Sci* **95:** 15281–15286.

Mosher RA, Schwach F, Studholme D, Baulcombe DC. 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci* **105:** 3145–3150.

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120:** 613–622.

Orlicky SM, Tran PT, Sayre MH, Edwards AM. 2001. Dissociable Rpb4–Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J Biol Chem* **276:** 10097–10102.

Paule MR, White RJ. 2000. Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Res* **28:** 1283–1298.

Pikaard CS, Haag JR, Ream T, Wierzbicki AT. 2008. Roles of RNA polymerase IV in gene silencing. *Trends Plant Sci* **13:** 390–397.

Pillai B, Sampath V, Sharma N, Sadhale P. 2001. Rpb4, a non-essential subunit of core RNA polymerase II of *Saccharomyces cerevisiae* is important for activated transcription of a subset of genes. *J Biol Chem* **276:** 30641–30647.

Pontes O, Costa-Nunes P, Vithayathil P, Pikaard CS. 2009. RNA polymerase V functions in *Arabidopsis* interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. *Mol Plant* **2:** 700–710.

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, Lagrange T. 2005. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev* **19:** 2030–2040.

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard C.S. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33:** 192–203.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.

Runner VM, Podolny V, Buratowski S. 2008. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3′ processing factors. *Mol Cell Biol* **28:** 1883–1891.

Russell J, Zomerdijk JC. 2006. The RNA polymerase I transcription machinery. *Biochem Soc Symp* **2006:** 203–216.

Saez-Vasquez J, Pikaard CS. 1997. Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc Natl Acad Sci* **94:** 11869–11874.

Schramm L, Hernandez N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16:** 2593–2620.

Sentenac A. 1985. Eukaryotic RNA polymerases. *Crit Rev Biochem* **18:** 31–91.

Sidorenko L, Chandler V. 2008. RNA-dependent RNA polymerase is required for enhancer-mediated transcriptional silencing associated with paramutation at the maize *p1* gene. *Genetics* **180:** 1983–1993.

Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL. 2009. A dominant mutation in *mediator of paramutation2*, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. *PLoS Genet* **5:** e1000725. doi: 10.1371/journal.pgen.1000725.

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

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57:** 758–771.

Ujvari A, Luse DS. 2006. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct*

Downloaded from symposium.cshlp.org on April 21, 2011 - Published by Cold Spring Harbor Laboratory Press

Mol Biol **13:** 49–54.

Verma-Gaur J, Rao SN, Taya T, Sadhale P. 2008. Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in *Saccharomyces cerevisiae*, reveals its involvement in transcription elongation. *Eukaryot Cell* **7:** 1009–1018.

Wei W, Dorjsuren D, Lin Y, Qin W, Nomura T, Hayashi N, Murakami S. 2001. Direct interaction between the subunit RAP30 of transcription factor IIF (TFIIF) and RNA polymerase subunit 5, which contributes to the association between TFIIF and RNA polymerase II. *J Biol Chem* **276:** 12266–12273.

Werner F. 2007. Structure and function of archaeal RNA polymerases. *Mol Microbiol* **65:** 1395–1404.

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135:** 635–648.

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41:** 630–634.

Woychik NA, Liao SM, Kolodziej PA, Young RA. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* **4:** 313–323.

Zaratiegui M, Irvine DV, Martienssen RA. 2007. Noncoding RNAs and gene silencing. *Cell* **128:** 763–776.

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE. 2007. Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci* **104:** 4536–4541.

Zheng B, Wang Z, Li S, Yu B, Liu JY, Chen X. 2009. Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in *Arabidopsis*. *Genes Dev* **23:** 2850–2860.