Cold Spring Harbor Symposia on Quantitative Biology

Darwin's "Abominable Mystery": The Role of RNA Interference in the Evolution of Flowering Plants

A. Cibrián-Jaramillo and R.A. Martienssen

Cold Spring Harb Symp Quant Biol 2009 74: 267-273 originally published online May 27, 2010
Access the most recent version at doi:10.1101/sqb.2009.74.051

Darwin's "Abominable Mystery": The Role of RNA Interference in the Evolution of Flowering Plants

A. Cibrián-Jaramillo[1,2,4] and R.A. Martienssen[3,4]

*[1]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York 10024; [2]The New York Botanical Garden, Bronx, New York 10458; [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724; [4]The New York Plant Genomics Consortium, Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York 10003**

Correspondence: martiens@cshl.edu

Darwin was famously concerned that the sudden appearance and rapid diversification of flowering plants in the mid-Cretaceous could not have occurred by gradual change. Here, we review our attempts to resolve the relationships among the major seed plant groups, i.e., cycads, ginkgo, conifers, gnetophytes, and flowering plants, and to provide a pipeline in which these relationships can be used as a platform for identifying genes of functional importance in plant diversification. Using complete gene sets and unigenes from 16 plant species, genes with positive partitioned Bremer support at major nodes were used to identify overrepresented gene ontology (GO) terms. Posttranscriptional silencing via RNA interference (RNAi) was overrepresented at several major nodes, including between monocots and dicots during early angiosperm divergence. One of these genes, *RNA-dependent RNA polymerase 6*, is required for the biogenesis of *trans*-acting small interfering RNA (tasiRNA), confers heteroblasty and organ polarity, and restricts maternal specification of the germline. Processing of small RNA and transfer between neighboring cells underlies these roles and may have contributed to distinct mutant phenotypes in plants, and in particular in the early split of the monocots and eudicots.

In an 1879 letter to J.D. Hooker, Darwin described the sudden appearance and rapid diversification of flowering plants in the mid Cretaceous as an "abominable mystery." Indeed, angiosperm diversification patterns presented an exception to his notion that nature evolves gradually *natura non facit saltum* (Friedman 2009). The key evolutionary processes that enable plants to adapt and diversify are only partially understood and mostly at the level of species, not at the level of major nodes, and less at the level of the evolving genome. On the 150th anniversary of *On the Origin of Species* (Darwin 1859), Darwin's mystery remains a fundamental issue in plant evolutionary biology not only within the angiosperms, but also within the gymnosperms. Namely, what are the fundamental evolutionary processes that enable plant species to adapt and diversify? One approach to understanding the origin and diversification of angiosperms and other seed plants is to identify genes or sets of genes that were critical in the divergence of key branches in plant evolution. By knowing gene function in at least some extant species, it should then be possible to correlate functional processes of interest with key steps such as the transition of plants from water to land or the evolution of the seed.

In principle, genes that were functionally important for branch divergence and plant diversification will have a phylogenetic signal that we can measure when we reconstruct phylogenies, through analysis of their effect on tree topology and branch support. Phylogenetic incongruence between a partitioned functional class of genes and the organismal phylogeny would suggest that the partition has experienced a unique evolutionary history relative to the organisms involved. In this way, incongruence of a particular class of genes in a partitioned analysis allows us to establish hypotheses about the evolution and potential function of these gene classes. Here, we use congruence measures of character evolution to mine genomes for patterns of protein function. In particular, we use modified elaborations of Bremer support—partitioned branch support (PBS) (Baker and DeSalle 1997), and partitioned hidden branch support (PHBS) (Gatesy et al. 1999). These measures can be used to evaluate the overall contribution (positive, negative, or neutral) of a particular gene to the various nodes or branches in a phylogenetic hypothesis. If one assumes that the tree obtained best represents the evolutionary history of the taxa involved, partitions that are in agreement or in conflict with the overall evolutionary history of the groups in the analysis can be detected and used to explain some of the more interesting organismal differences among taxa. This phylogenomic framework is powerful because it integrates experimental and genomic data to enable predictions of gene function, allowing us to tease apart the role of evolutionary change in protein function (Eisen 1998; Eisen and Fraser 2003; Sjölander 2004; Brown and Sjölander 2006).

*Members of the Consortium are Gloria Coruzzi (New York University), Rob DeSalle (American Museum of Natural History), Dennis Stevenson (The New York Botanical Garden), W. Richard McCombie (Cold Spring Harbor Laboratory), Ernest Lee (American Museum of Natural History), Sergios-Orestis Kolokotronis (American Museum of Natural History), Manpreet Katari (New York University), A.C.J. (American Museum of Natural History/The New York Botanical Garden), A.C.-J. (The New York Botanical Gardens), and R.M. (Cold Spring Harbor Laboratory).

Plant phylogenies to date are based on a few nuclear and plastid markers (for review, see APGIII 2009; Mathews 2009). By making use of the increasingly available genomic and expressed sequence tag (EST) data, as well as in-house data provided by The New York Plant Genomics Consortium, it was possible to construct a phylogeny of protein sequences from 2557 orthologous genes spanning 16 plant species. Species were chosen as representatives of major groups of angiosperms, gymnosperms, and nonseed plants, and measures of support were used to identify proteins and characters that may have functional significance across those 16 taxa (Cibrián-Jaramillo et al. 2010).

Our phylogenomic approach allows all character information to interact freely and reveal a more accurate description of species relationships, and at the same time, it makes it possible to observe snapshots of how genes or groups of genes may have evolved in the context of the overall phylogeny. These sets of genes themselves are a hypothesis, and their relevance to that node can be tested further based on measures of selection and explicit experimental analyses. It is clear that genome-level sequencing and large EST studies are rapidly growing, expanding the number of gene partitions and ways of partitioning phylogenetic information that are available. Our platform can easily incorporate the information simultaneously, contributing to the efficient integration of genomic and experimental data and enlightening the evolutionary processes driving plant diversification (Chiu et al. 2006; Cibrián-Jaramillo et al. 2010).

METHODS

Expanding on the analysis of De la Torre-Bárcena et al. (2009), we assembled a matrix of all available genomic and EST data to date for 16 plant species including 11 seed plants—five angiosperms (*Amborella*, rice, *Arabidopsis*, poplar, and grape) and six gymnosperms (*Cryptomeria*, pine, two cycads, gingko, *Gnetum*, and *Welwitschia*)—and four seed-free plants—Filicalian fern (*Adiantum*), a thalloid liverwort (*Marchantia*), a moss (*Physcomitrella*), and a Lycophyte (*Selaginella*) (Table 1).

Orthology of genes was established using OrthologID (Chiu et al. 2006), http://nypg.bio.nyu.edu/orthologid. OrthologID is an automated approach to sort query sequences into gene family membership and determine sets of orthologs from the gene trees. All ortholog groups reflecting coding genes are then assembled into a concatenated matrix of 1,062,841 amino acids representing 2557 proteins (genes), with delineated data partitions for each gene (for other methodological details, see Cibrián-Jaramillo et al. 2010). A maximum parsimony tree was generated using all concatenated genes in a simultaneous analysis (SA) and individually (partitioned data). Parsimony analysis was performed in PAUP* 4b10 using equal weights (Swofford 2003). Branch support was evaluated using the nonparametric bootstrap (2000 replicates) and jackknife (50% and 30% removal) methods in PAUP (Felsenstein 1985; Farris et al. 1996).

Once the most parsimonious tree is identified through character congruence, we can examine the partitions to say something about their function. The delineation of data partitions allows the contribution of a gene (partition) to a branch to be assessed using congruence measures of support. We used a customized Perl script to calculate individual tree statistics including PBS and PHBS (Cibrián-Jaramillo et al. 2010). By definition, for a particular combined data set, a particular node (branch), and a particular data partition, PBS is the minimum number of character steps for that partition on the shortest topologies for the combined data set which do not contain that node, minus the minimum number of character steps for that partition on the shortest topologies for the combined data set that do contain that node (Baker and DeSalle 1997). PHBS is the difference between PBS for that data partition and the Bremer support value (Bremer 1988, 1994) for that node for that data partition (Gatesy et al. 1999). Values for these metrics can be positive, zero, or negative, and the value can indicate the direction of support for the overall concatenated hypothesis: Positive lends support, zero is neutral, and negative gives conflicting support (Gatesy et al. 1999).

A gene ontology (GO) term was established for each gene based on orthology with an *Arabidopsis* gene ID number using the current TAIR v8 database (http://www.arabidopsis.org). To determine which branch contains enrichment of a certain molecular or biological function, statistically overrepresented GO categories at each partition were compared to the distribution of that GO term in the *Arabidopsis* genome (considered a "baseline distribution"; Cibrián-Jaramillo et al. 2010). Because each branch is composed of partitions which represent genes that provide positive, negative, or neutral support, genes were first grouped into four sets: (1) genes that had a positive value for PBS (apparent), (2) genes that had a positive value for PHBS (hidden) support, (3) genes with neutral PBS, and (4) genes with neutral (zero) PHBS (no evolutionary signature for each branch).

Sungear (Poultney et al. 2007) implemented in Virtual Plant (http://www.virtualplant.org) was used to compare different sets of gene lists against *Arabidopsis*. GO term

Table 1. List of Species and Genomic Sources

Species	Genomic database
Adiantum capillus-veneris	TIGR PlantTA
Amborella trichopoda	TIGR PlantTA
Arabidopsis thaliana[a]	TAIR
Cryptomeria japonica	TIGR PlantTA
Cycas rumphii	CSHL/TIGR PlantTA
Ginkgo biloba	CSHL/TIGR PlantTA
Gnetum gnemon	CSHL/TIGR PlantTA
Marchantia polymorpha	JCVI
Oryza sativa[a]	JGI
Pinus taeda	TIGR PlantTA
Populus trichocarpa[a]	JGI
Selaginella moellendorffii	TIGR PlantTA
Vitis vinifera[a]	Genoscope
Welwitschia mirabilis	TIGR PlantTA
Zamia fischeri	CSHL/TIGR PlantTA

[a]Complete genomes: (TIGR) http://compbio.dfci.harvard.edu/tgi/plant.html; (PlantTA) http://plantta.jcvi.org; (CSHL) http://www.cshl.edu; (JCVI) http://www.jcvi.org; (JGI) http://www.jgi.doe.gov; (Genoscope) http://www.genoscope.cns.fr/spip.

overrepresentation is measured by a *z*-score representing the number of standard deviations by which a particular observation (i.e., number of genes) is above or below the mean (Dudoit et al. 2004; Gutiérrez et al. 2007). Partitions with overrepresented GO terms and positive PBS within the angiosperms, nodes 4 through 7, were further investigated using Biomaps (Wang et al. 2004) as implemented in VirtualPlant. This tool provides a different measure of overrepresentation by using a hypergeometric distribution and significance based on a *p*-value ($p < 0.05$). Biomaps was used to compare the observed distribution of genes at each branch to the distribution of those GOs terms associated with *Arabidopsis* genes found in the matrix.

RESULTS

The resulting tree is identical in topology to a tree previously obtained with maximum parsimony (MP) and maximum likelihood (ML), although with fewer partitions (1200) and various combinations of ingroup and outgroup taxa (De la Torre-Bárcena et al. 2009) (Fig. 1). Other tree manipulations and details regarding phylogenetic analyses are summarized in De la Torre-Bárcena et al. (2009) and Cibrián-Jaramillo et al. (2010). A subset of *Arabidopsis* orthologs 1503 (58.7%) had at least one functional GO category (the total number of GO categories matched is 1872). The overall GO term distribution of the matrix had no significant biases (that would suggest a methodological bias) compared to *Arabidopsis* (Cibrián-Jaramillo et al. 2010).

A number of genes were found belonging to GO categories with very low probabilities of occurring by chance at the observed frequencies (*z*-scores) for both positive PBS and PHBS, with no significant outliers with neutral PHBS or PBS. Positive PHBS genes provide additional support at a particular node in the simultaneous analysis of all data partitions. Figure 2 illustrates the distribution of sets of overrepresented genes (represented by vessels) for PHBS values at each node in the tree. Node 6 (*Arabidopsis*, *Populus, Vitis*) and node 7 (*Populus*, *Vitis*) have the largest outlier vessels.

Within the angiosperms (Biomaps), overrepresented GO terms with both PBS and PHBS support included photosynthesis, development, and hormone-related functional categories (Figs. 1 and 2). A functional group was of exceptional interest: genes involved in posttranscriptional gene silencing, in particular *AGO1* and *RDR6* within the rosids (*Arabidopsis*, *Populus*, *Vitis*). Notably, character comparison for *AGO1* (not shown) and *RDR6* (Fig. 3) revealed a number of amino acid substitutions at regions in proteins with known functional importance (Marchler-Bauer et al. 2007). For *RDR6*, the *SHOOT LESS2* gene (*SHL2*) is the rice ortholog of *RDR6* in *Arabidopsis*. The *shl2-10* allele, *shl2*, has a G614D mutation, responsible for that mutant phenotype (Nagasaki et al. 2007). This specific site is one of those supporting cladogenetic variations in our matrix (Fig. 3), providing positive branch support as apomorphic for monocots (Cibrián-Jaramillo et al. 2010).

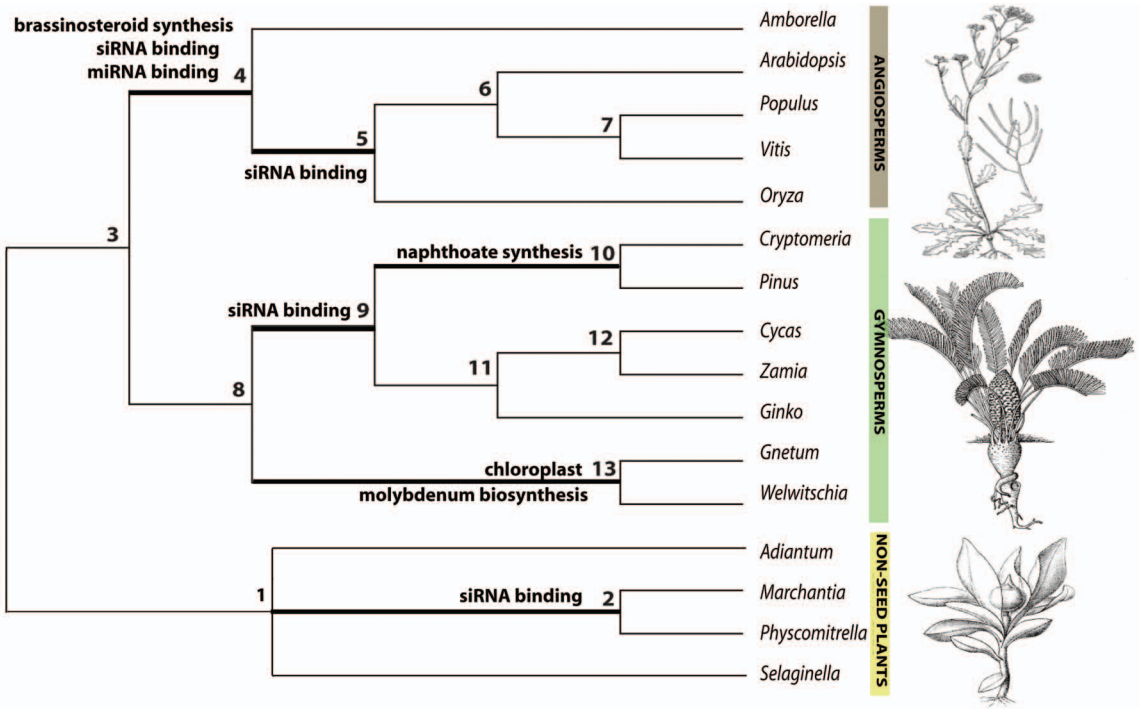

Figure 1. Phylogenetic relationships of seed plants using 2557 proteins inferred with maximum parsimony. All nodes showed bootstrap (2000 replicates) and jackknife (1000 replicates) support values above 99%. Overrepresented GOs with the most important functional categories are shown at the base of the nodes. (Modified, with permission, from Cibrián-Jaramillo et al. 2010 [© Oxford University Press].)

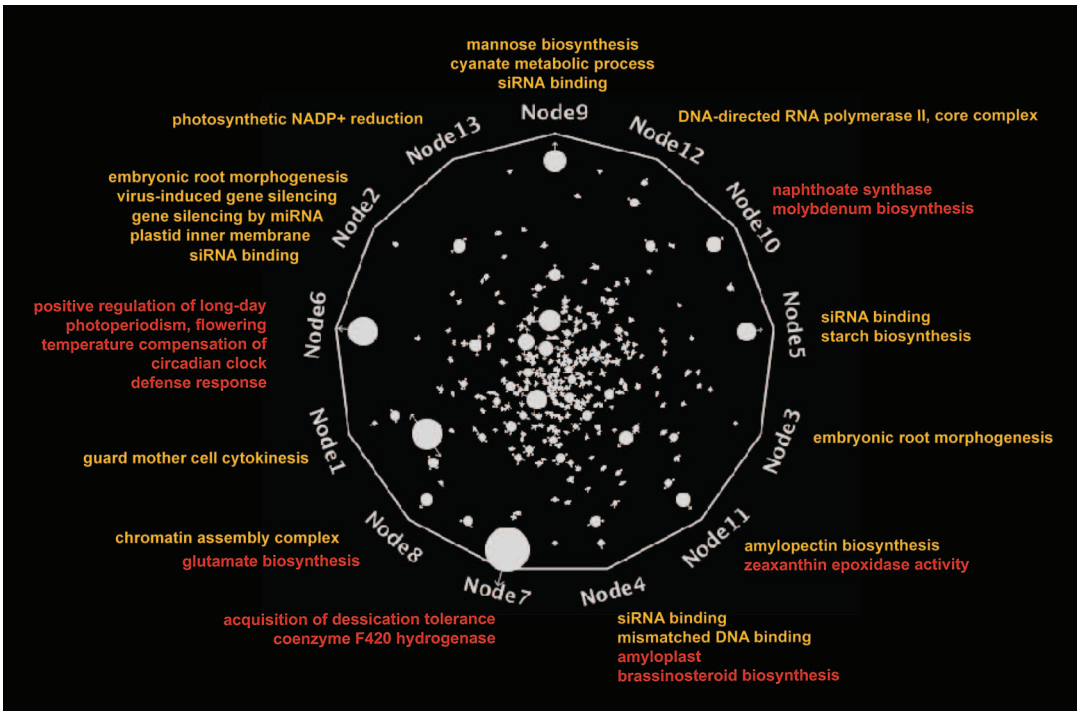

Figure 2. Distribution of genes across nodes. Sungear allows for the visual and statistical analysis of overlapping relationships among different lists of data and Boolean combinations. Each polygon corresponds to a particular node in the phylogeny. The circles with arrows within the polygon are called vessels, which represent genes with a positive *z*-score, from the set of categories with positive PHBS values. The position and the arrows of the vessels identify which node the genes are from, and the size of the vessel is relative to the number of genes in that vessel. The most interesting gene categories are written on each node (red categories are only found in that node, whereas yellow categories are shared across various nodes).

DISCUSSION

The topology described here recovers major groups of seed plants as all previous morphological analyses and most molecular analyses with monophyletic seed plants have done: the cycads, the conifers, the gnetophytes, and the angiosperms. We support the gnetophytes as the sister group to all other gymnosperms, congruent with phylogenetic studies using phytochrome genes (Mathews and Donoghue 2000; Schmidt and Schneider-Poetsch 2002; Mathews 2009), *AGAMOUS*-like genes (Winter et al. 1999; Becker et al. 2003), and *FLORICAULA/LEAFY* (Frohlich and Parker 2000). These results are to a great extent congruent with the angiosperm phylogeny group (APG III) (APGIII 2009). A more detailed explanation of the importance of a simultaneous analysis, as well as of support dynamics, the role of outgroup choice, taxon sampling, and missing data is presented elsewhere (De la Torre-Bárcena et al. 2009; Cibrián-Jaramillo et al. 2010).

Genes with Evolutionary and Functional Relevance

Overrepresented functional categories that are common throughout nodes are largely metabolic processes, such as photosynthesis. This distribution is concordant with their importance in key biochemical pathways that plants have developed in response to major environmental stress. For example, changes in photosynthetic chemical pathways are used not only to adapt to novel light conditions, but also to reduce evaporative water loss that was probably required from the transition from water to land and when plants colonized new environments (Bohnert et al. 1988). Interestingly, the Gnetophyta (node 13) had the highest number of overrepresented photosynthetic genes (Cibrián-Jaramillo et al. 2010). The gnetophyte *Welwitschia mirabilis* has a crassulacean acid metabolism (CAM) photosynthetic pathway in which stomata are open at night, avoiding water diffusion during the day (von Willert et al. 2005). Another member, *Ephedra*, is found in semiarid to desert conditions exposed to water stress during part of the year. Most *Gnetum* species are distributed in lowland tropical rainforests and are uniquely characterized by a relatively lower photosynthetic capacity as well as reduced capacity for stem water transport (Feild and Balun 2008).

At the other end of the spectrum are overrepresented gene categories directly related to specific traits or phenotype characteristics of that clade. For example, overrepresented amylopectin genes at the conifers (node 9) and mannose biosynthesis genes at the cycad node (node 11) may have a direct association to their morphology (Fig. 2). Amylopectin is fundamental to the manoxylic wood in cycads, and it differs from the pycnoxylic wood in conifers and the Gnetales, in which mannose is an important component (Greguss 1955).

Within the angiosperms, plant hormones and genes involved in circadian clock and photoperiodism were among the most interesting overrepresented partitions.

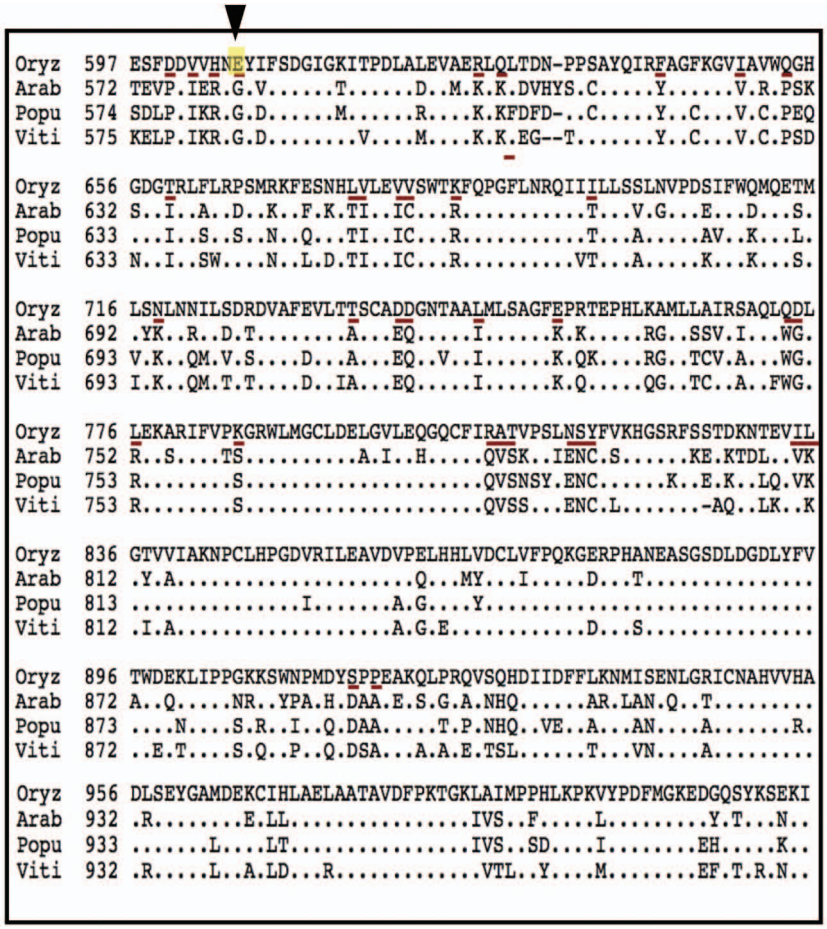

Figure 3. Character comparison for *RDR6* among angiosperms reveals amino acid substitutions at regions in proteins with known functional importance. Shown is part of the alignment in our matrix that corresponds to the *RDR6* domain. The *SHOOTLESS2* gene (*SHL2*) is the rice ortholog of *RDR6* (from *Arabidopsis*). In the *shl2-10* allele, *shl2* has a G614D mutation, responsible for the mutant phenotype, i.e., functionally important site (Nagasaki et al. 2007). This site is one of those supporting cladogenetic variations in our matrix, i.e., providing positive branch support for the split between monocots and the rest of the angiosperms. Substitutions unique to rice throughout the domain are underlined in red. Approximations of domain span are based on Marchler-Bauer et al. (2007). (Modified, with permission, from Cibrián-Jaramillo et al. 2010 [© Oxford University Press].)

Brassinosteroid genes were found to be uniquely overrepresented in the angiosperm clade (node 4) (Fig. 2). Brassinosteroid hormones differ in their signaling from other hormones, with a relatively longer pathway than either auxin or gibberellin (Bajguz and Tretyn 2003). Carotenoid biosynthesis factors, involved in shoot-branching and long-range signaling (Mouchel and Leyser 2007), were identified in the same node. Genes that are involved in the regulation of the circadian clock, photoperiodism, and growth habit (Balasubramanian et al. 2006) are overrepresented in the rosids (node 6). Patterns of hormone expression and regulation of circadian clock, and their specific function in the rest of the angiosperms, must be tested in future molecular and developmental studies, but their relevance is highlighted here.

RNAi in Plant Evolution

Genes involved in posttranscriptional regulation by small RNAs are highly overrepresented functional categories in both gymnosperms and angiosperms (Fig. 1). They have among the highest significance values for overrepresented genes in the split between *Amborella* and the rest of the angiosperms (node 4) and in the split between monocots and eudicots (node 5). microRNAs (miRNAs) and small interfering RNAs (siRNAs) are important for developmental phenotypes (Willman and Poethig 2005; Sunkar and Zhu 2007; Wang et al. 2007), and some are highly conserved.

Two genes in particular, *Argonaute* (*AGO1*) and *RNA-dependent RNA polymerase 6 (RDR6*), are critical in developmental aspects of RNAi. They have roles in various stages of embryo and leaf development, polarity, and shape through *trans*-acting siRNA and miRNA pathways (Kidner and Martienssen 2004, 2005; Peragine et al. 2004). *AGO1* provides positive phylogenetic support for the angiosperm clade (node 4), for the split of *Amborella* and the rest of the angiosperms (node 5), and for the eudicot clade only (node 6). *RDR6* provides support for the split of the eudicots from *Amborella* and rice (node 6).

Their overrepresentation and phylogenetic contribution are highly relevant given character analysis in our phylogeny. In particular, we found a number of amino acid substitutions among species in the clades with high support, at regions in proteins with known functional importance. For *AGO1*, mutations unique to rice were found in the PAZ nucleic-acid-binding interface and in regions that correspond to the 5′ guide strand anchoring domain and the PIWI catalytically active domain. Interestingly, mutants in *RDR6*, which support the dicot clade, have much milder phenotypes in *Arabidopsis* (Adenot et al. 2006; Fahlgren et al. 2006; Garcia et al. 2006) than in the monocot rice (Nagasaki et al. 2007). Asymmetry and shoot meristem organization in the monocotyledonous embryo of rice are strongly affected, whereas the dicotyledonous embryo of *Arabidopsis* remains radically symmetric and germinates normally. In terms of target gene expression, rice *rdr6* mutant embryos lose some of their asymmetry, resembling dicotyledonous embryos in this respect, although profound differences remain. Unique changes to rice are sites of potentially important mutants. Overall, *AGO1* and *RDR6* (and then, the processing or transport of *trans*-acting siRNA) are implicated in this defining feature of the angiosperm seed. Recently, *RDR6* has also been implicated in small-RNA-mediated suppression of gamete formation in the *Arabidopsis* ovule, a phenotype related to asexual seed formation, or apomixis (Olmedo-Monfil et al. 2010). Expression analysis using these sequence variants will confirm a role for these amino acid residues in determining significant phenotypic effects of ecological and evolutionary importance.

CONCLUSIONS

We demonstrate a novel method using phylogenomic tools to postulate hypotheses of gene function in the evolution of major plant groups. Functional hypotheses can be further coupled with expression and genetic data to arrive at better gene annotations and functional analyses for genome-level studies. Our findings help to guide plant ecological genomics studies and enlighten the precise evolutionary mechanisms driving the diversification of plant species, helping to gradually unravel Darwin's abiding and perplexing mystery.

ACKNOWLEDGMENTS

We are grateful to our colleagues in the New York Plant Genomics (NYPG) Consortium whose work we review here, especially Ernest Lee, Sergios Kolokotronis, and Dennis Stevenson. A.C.J. is funded by the Lewis B. and Dorothy Cullman Fellowship at the American Museum of Natural History and The New York Botanical Garden. The NYPG Consortium is supported by National Science Foundation grant 0421604 "Genomics of Comparative Seed Evolution" to Gloria Coruzzi (New York University), Rob DeSalle (American Museum of Natural History), Dennis Stevenson (The New York Botanical Garden), Dick McCombie (Cold Spring Harbor Laboratory) and R.M. (Cold Spring Harbor Laboratory).

REFERENCES

Adenot X, Elmayan T, Lauressergues D, Boutet S, Bouché N, Gasciolli V, Vaucheret H. 2006. DRB4-dependent *TAS3 trans*-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16:** 927–932.

APGIII. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* **161:** 105–121.

Bajguz A, Tretyn A. 2003. The chemical characteristic and distribution of brassinosteroids in plants. *Phytochemistry* **62:** 1027–1046.

Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst Biol* **46:** 654–673.

Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann JC, Weigel D. 2006. The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. *Nat Genet* **38:** 711–715.

Becker A, Saedler H, Theissen G. 2003. Distinct MADS-box gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev Genes Evol* **213:** 567–572.

Bohnert HJ, Ostrem JA, Cushman JC, Michalowski CB, Rickers J, Meyer G, deRocher EJ, Vernon DM, Krueger M, Vazquez-Moreno L, et al. 1988. *Mesembryanthemum crystallinum*, a higher plant model for the study of environmentally induced changes in gene expression. *Plant Mol Biol Rep* **6:** 10–28.

Bremer K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795–803.

Bremer K. 1994. Branch support and tree stability. *Cladistics* **10:** 295–304.

Brown D, Sjölander K. 2006. Functional classification using phylogenomic inference. *PLoS Comput Biol* **2:** e77.

Chiu JC, Lee EK, Egan MG, Sarkar IN, Coruzzi GM, DeSalle R. 2006. OrthologID: Automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* **22:** 699–707.

Cibrián-Jaramillo A, De la Torre-Bárcena JE, Lee KE, Katari MS, Little DP, Stevenson DW, Martienssen R, Coruzzi G, DeSalle R. 2010. Using phylogenomic patterns and gene ontology to identify proteins of importance in plant evolution. *Genome Biol Evol* (in press).

Darwin C. 1859. *On the origin of species by means of natural selection,* 1st ed. Murray, London.

De la Torre-Bárcena JE, Kolokotronis SO, Lee EK, Stevenson DW, Coruzzi GM, DeSalle R. 2009. The impact of outgroup choice and missing data on major seed plant phylogenetics using genome-wide EST data. *PLoS ONE* **4:** e5764.

Dudoit S, van der Laan MJ, Pollard KS. 2004. Multiple testing. I. Single-step procedures for control of general type I error rates. *Stat Appl Genet Mol Biol* **3:** 1–69.

Eisen JA. 1998. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* **8:** 163–167.

Eisen JA, Fraser CM. 2003. Phylogenomics: Intersection of evolution and genomics. *Science* **300:** 1706–1707.

Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, Alexander AL, Carrington J. 2006. Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr Biol* **9:** 939–944.

Farris J, Albert V, Källersjö M, Lipscomb D, Kluge A. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12:** 99–124.

Feild TS, Balun L. 2008. Xylem hydraulic and photosynthetic function of *Gnetum* (Gnetales) species from Papua New Guinea. *New Phytol* **177:** 665–675.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783–791.

Friedman WE. 2009. The meaning of Darwin's 'abominable mystery.' *Am J Bot* **96:** 5.

Frohlich MW, Parker DS. 2000. The mostly male theory of flower

evolutionary origins: From genes to fossils. *Syst Bot* **25:** 155–170.

Garcia D, Collier SA, Byrne ME, Martienssen RA. 2006. Specification of leaf polarity in *Arabidopsis* via the *trans*-acting siRNA pathway. *Curr Biol* **16:** 933–938.

Gatesy J, O'Grady P, Baker RH. 1999. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* **15:** 271–313.

Greguss P. 1955. *Identification of living gymnosperms on the basis of xylotomy* (transl. L Jocsik). Akademiai Kiado, Budapest.

Gutiérrez RA, Gifford ML, Poultney C, Wang R, Shasha DE, Coruzzi GM, Crawford NM. 2007. Insights into the genomic nitrate response using genetics and the Sungear Software System. *J Exp Bot* **58:** 2359–2367.

Kidner CA, Martienssen RA. 2004. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* **428:** 81–84.

Kidner CA, Martienssen RA. 2005. The role of ARGONAUTE1 (AGO1) in meristem formation and identity. *Dev Biol* **280:** 504–517.

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, et al. 2007. CDD: A conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35:** D237–D240.

Mathews S. 2009. Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. *Am J Bot* **96:** 228–236.

Mathews S, Donoghue MJ. 2000. Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. *Int J Plant Sci* **161:** 41–55.

Mouchel CF, Leyser O. 2007. Novel phytohormones involved in long-range signaling. *Curr Opin Plant Biol* **10:** 473–476.

Nagasaki H, Itoh J, Hayashi K, Hibara K, Satoh-Nagasawa N, Nosaka M, Mukouhata M, Ashikari M, Kitano H, Matsuoka M, et al. 2007. The small interfering RNA production pathway is required for shoot meristem initiation in rice. *Proc Natl Acad Sci* **104:** 14867–14871.

Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* **464:** 628–632.

Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. 2004. *SGS3* and *SGS2/SDE1/RDR6* are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev* **18:** 2368–2379.

Poultney CS, Gutiérrez RA, Katari MS, Gifford ML, Palen WB, Coruzzi GM, Shasha DE. 2007. Sungear: Interactive visualization and functional analysis of genomic datasets. *Bioinformatics* **23:** 259–261.

Schmidt M, Schneider-Poetsch HA. 2002. The evolution of gymnosperms redrawn by phytochrome genes: The Gnetatae appear at the base of the gymnosperms. *J Mol Evol* **54:** 715–724.

Sjölander K. 2004. Phylogenomic inference of protein molecular function: Advances and challenges. *Bioinformatics* **20:** 170–179.

Sunkar R, Zhu JK. 2007. Micro RNAs and short-interfering RNAs in plants. *J Integr Plant Biol* **49:** 817–826.

Swofford D. 2003. *PAUP*: Phylogenetic analysis using parsimony* (**and other methods*). Sinauer, Sunderland, MA.

von Willert DJ, Armbruster N, Drees T, Zaborowski M. 2005. *Welwitschia mirabilis:* CAM or not CAM—What is the answer? *Funct Plant Biol* **32:** 389–395.

Wang R, Tischner R, Gutiérrez RA, Hoffman M, Xing X, Chen M, Coruzzi G, Crawford NM. 2004. Genomic analysis of the nitrate response using a nitrate reductase-null mutant of *Arabidopsis*. *Plant Physiol* **136:** 2512–2522.

Wang Y, Stricker HM, Gou D, Liu L. 2007. MicroRNA: Past and present. *Front Biosci* **12:** 2316–2329.

Willman MR, Poethig RS. 2005. Time to grow up: The temporal role of small RNAs in plants. *Curr Opin Plant Biol* **8:** 548–552.

Winter KU, Becker A, Munster T, Kim JT, Saedler H, Theissen G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc Natl Acad Sci* **96:** 7342–7347.