Evidence for Network Evolution in an *Arabidopsis* Interactome Map

Arabidopsis Interactome Mapping Consortium

Science **333**, 601 (2011);
DOI: 10.1126/science.1203877

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this infomation is current as of August 2, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/333/6042/601.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2011/07/27/333.6042.601.DC1.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/content/333/6042/601.full.html#related

This article **cites 42 articles**, 15 of which can be accessed free:
http://www.sciencemag.org/content/333/6042/601.full.html#ref-list-1

This article has been **cited by** 2 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/333/6042/601.full.html#related-urls

This article appears in the following **subject collections:**
Botany
http://www.sciencemag.org/cgi/collection/botany

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2011 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

molecular MTI markers was abolished in the *fls2* mutant, which lacks the PRR receptor for flg22 peptide, and largely impaired in *pfd6-1* (fig. S13). These results link PFD6 to MTI downstream of FLS2 PRR receptor function (*10*, *33*). Collectively, these results (Fig. 4) validate the biological significance of PPIN-1 and confirm that pathogen effectors target host proteins that are required for effective defense or pathogen fitness. To facilitate further hypothesis testing, we present the local networks for the five significantly targeted hubs (Fig. 2D and table S4) and point out connections to cellular functions potentially relevant to immune system function (figs. S14 to S18).

Conclusions. Our analyses reveal that oomycete and bacterial effectors separated by ~2 billion years of evolution target an overlapping subset of plant proteins that include well-connected cellular hubs. Our functional validation supports the notion that effectors are likely to converge onto interconnected host machinery to suppress effective host defense and to facilitate pathogen fitness. We predict that many of the 165 effector targets we defined will also be targets of additional, independently evolved effectors from other plant pathogens. We anticipate that effectors that target highly connected cellular proteins fine-tune cellular networks to increase pathogen fitness and that evolutionary forces integrate appropriate immune responses with those perturbations. As proposed in the guard hypothesis, our data are consistent with indirect connections between pathogen effectors and NB-LRR immune receptors, at least for the NB-LRR fragments represented in PPIN-1. The high degree of the effector targets argues against a decoy role for these proteins. Although the concept of cellular decoys evolved to intercept pathogen effectors is attractive, and likely true in one case in the plant immune system (*3*), these are expected to have few, if any, additional cellular functions and, as such, would likely have fewer interaction partners in the protein interaction network. Most of the 673 immune interactors have no previously described immune-system function. Our results bridge plant immunology, which predicted that effectors should target common proteins, and network science, which proposes that hubs should be targets for network manipulation (*25*–*28*). Derivation of general rules regarding the organization and function of host cellular machinery required for effective defense against microbial infection, as well as detailed mechanistic understanding of how pathogen effectors manipulate these machines to increase their fitness, will facilitate improvement of plant immune system function.

References and Notes

1. C. Zipfel, *Curr. Opin. Plant Biol.* **12**, 414 (2009).
2. T. Boller, S. Y. He, *Science* **324**, 742 (2009).
3. P. N. Dodds, J. P. Rathjen, *Nat. Rev. Genet.* **11**, 539 (2010).
4. J. L. Dangl, J. D. Jones, *Nature* **411**, 826 (2001).
5. J. D. Jones, J. L. Dangl, *Nature* **444**, 323 (2006).
6. E. Lukasik, F. L. Takken, *Curr. Opin. Plant Biol.* **12**, 427 (2009).
7. G. van Ooijen *et al.*, *J. Exp. Bot.* **59**, 1383 (2008).
8. D. A. Baltrus *et al.*, *PLoS Pathog.* **7**, e1002132 (2011).
9. L. Baxter *et al.*, *Science* **330**, 1549 (2010).
10. Glossary, materials and methods, supporting figures, and supporting tables are available as supporting material on *Science* Online.
11. Arabidopsis Interactome Mapping Consortium, *Science* **333**, 601 (2011).
12. M. Dreze *et al.*, *Methods Enzymol.* **470**, 281 (2010).
13. P. Braun *et al.*, *Nat. Methods* **6**, 91 (2009).
14. M. E. Cusick *et al.*, *Nat. Methods* **6**, 39 (2009).
15. H. Yu *et al.*, *Science* **322**, 104 (2008).
16. J. D. Lewis, D. S. Guttman, D. Desveaux, *Semin. Cell Dev. Biol.* **20**, 1055 (2009).
17. Z. Y. Peng *et al.*, *Nucleic Acids Res.* **37**, (Database issue), D975 (2009).
18. X. Tan *et al.*, *BMC Plant Biol.* **7**, 56 (2007).
19. C. Zipfel *et al.*, *Nature* **428**, 764 (2004).
20. T. B. Sackton *et al.*, *Nat. Genet.* **39**, 1461 (2007).
21. E. B. Holub, *Nat. Rev. Genet.* **2**, 516 (2001).
22. P. Zhang *et al.*, *Plant Physiol.* **138**, 27 (2005).
23. Y. Jaillais, J. Chory, *Nat. Struct. Mol. Biol.* **17**, 642 (2010).
24. R. Albert, H. Jeong, A. L. Barabasi, *Nature* **406**, 378 (2000).
25. B. de Chassey *et al.*, *Mol. Syst. Biol.* **4**, 230 (2008).
26. M. D. Dyer *et al.*, *PLoS ONE* **5**, e12089 (2010).
27. M. A. Calderwood *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7606 (2007).
28. P. Uetz *et al.*, *Science* **311**, 239 (2006).
29. J. M. Alonso *et al.*, *Science* **301**, 653 (2003).
30. A. Sessions *et al.*, *Plant Cell* **14**, 2985 (2002).
31. R. Lozano-Durán *et al.*, *Plant Cell* **23**, 1014 (2011).
32. G. Gusmaroli, P. Figueroa, G. Serino, X. W. Deng, *Plant Cell* **19**, 564 (2007).
33. S. Robatzek, D. Chinchilla, T. Boller, *Genes Dev.* **20**, 537 (2006).
34. K. Tsuda, M. Sato, J. Glazebrook, J. D. Cohen, F. Katagiri, *Plant J.* **53**, 763 (2008).

Acknowledgments: This work was funded by NIH GM-066025, NSF 2010 0929410, and U.S. Department of Energy (DOE) FG02-95ER20187 to J.L.D.; U.K. Biotechnology and Biological Sciences Research Council E024815, F005806 and G015066 to J.B.; NSF 0703905 to M.V., J.R.E., D.E.H.; NIH P50-HG004233 to M.V.; and NSF 0520253, 0313578 and 0726408 to J.R.E. D.M. was supported by AGRONOMICS LSHG-CT-2006-037704 from Sixth Framework Programme of the European Commission to C. Lurin. We acknowledge that the NSF funded the *Arabidopsis* Biological Research Center and the Salk Institute Genomic Analysis Laboratory (SIGnAL) projects for seeds and clones, respectively. We thank L. Baxter (Warwick Systems Biology, UK) for *Arabidopsis*/Papaya ortholog identification; B. Charloteaux (Center of Cancer Systems Biology, Boston, USA) for assisting in some bioinformatics analyses; B. Kemmerling (University of Tuebingen, Germany) for several RLK clones not contained in SIGnAL; and C. Somerville and Y. Gu (University of California, Berkeley, USA), T. Mengiste (Purdue University, USA), and X.-W. Deng (Yale University, USA) for seeds. The EU Effectoromics Consortium was funded by the European Research Area in Plant Genomics and includes A. Cabral and G. van den Ackerveken (Utrecht University, The Netherlands); J. Bator, R. Yatusevich, S. Katou and J. Parker (Max Planck Institute for Plant Breeding Research, Cologne, Germany); G. Fabro and J. Jones (The Sainsbury Laboratory, Norwich, UK); and M. Coates and T. Payne (University of Warwick, Warwick, UK). M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation, Belgium). Binary interaction data are supplied in table S2 in SOM. Homozygous mutant seed stocks noted in Figure 4 are available from the Arabidopsis Biological Resource Center (ABRC). Author contributions are listed in the SOM.

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6042/596/DC1
Materials and Methods
SOM Text
Figs. S1 to S18
Tables S1 to S10
References

1 February 2011; accepted 6 June 2011
10.1126/science.1203659

Evidence for Network Evolution in an *Arabidopsis* Interactome Map

Arabidopsis **Interactome Mapping Consortium***†

Plants have unique features that evolved in response to their environments and ecosystems. A full account of the complex cellular networks that underlie plant-specific functions is still missing. We describe a proteome-wide binary protein-protein interaction map for the interactome network of the plant *Arabidopsis thaliana* containing about 6200 highly reliable interactions between about 2700 proteins. A global organization of plant biological processes emerges from community analyses of the resulting network, together with large numbers of novel hypothetical functional links between proteins and pathways. We observe a dynamic rewiring of interactions following gene duplication events, providing evidence for a model of evolution acting upon interactome networks. This and future plant interactome maps should facilitate systems approaches to better understand plant biology and improve crops.

Classical genetic and molecular approaches have provided fundamental understanding of processes such as growth control or development and molecular descriptions of genotype-to-phenotype relationships for a variety of plant systems. Yet, more than 60% of the protein-coding genes of the model plant *Arabidopsis thaliana* (hereafter *Arabidopsis*) remain functionally uncharacterized. Knowledge about the biological organization of macromolecules in complex and dynamic "interactome" networks is lacking for *Arabidopsis* (fig. S1 and tables S1 and S2), depriving us of an understanding of how genotype-to-phenotype relationships are mediated at the systems level (*1*).

---

*All authors with their affiliations and contributions are listed at the end of the paper.

†To whom correspondence should be addressed. E-mail: marc_vidal@dfci.harvard.edu; ecker@salk.edu; pascal_braun@dfci.harvard.edu; david_hill@dfci.harvard.edu

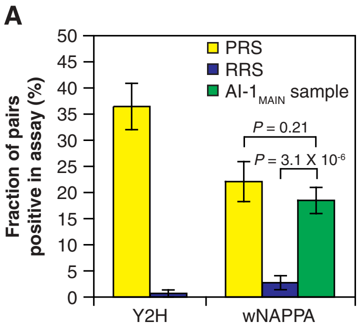
A high-quality binary protein-protein interactome map for *Arabidopsis.* To generate a map of the *Arabidopsis* interactome network, we used a collection of ~8000 open reading frames representing ~30% of its predicted protein-coding genes (fig. S2 and table S3) (*2*, *3*). We tested all pairwise combinations of proteins encoded by these constructs (space 1) with an improved high-throughput binary interactome mapping pipeline based on the yeast two-hybrid (Y2H) system (fig. S2) (*3*, *4*). Confirmed pairs were assembled into a data set of 5664 binary interactions between 2661 proteins, called *Arabidopsis* Interactome version 1 "main screen" (AI-1$_{MAIN}$) (table S4).

The quality of AI-1$_{MAIN}$ was evaluated against a positive reference set (PRS) of 118 well-documented, manually recurated (*5*) *Arabidopsis* protein-protein interactions and a random reference set (RRS) of 146 random protein pairs (fig. S3 and table S5) (*3*, *5*–*9*). We determined the fraction of true biophysical interactions in AI-1$_{MAIN}$, its precision, to be ~80%, by comparing the validation rates of a random sample of 249 interactions from AI-1$_{MAIN}$ to those of the PRS and RRS in a well–nucleic acid programmable protein array (wNAPPA) protein-protein interaction assay (Fig. 1A, fig. S4, and table S5) (*3*, *8*).

To estimate the size of the complete *Arabidopsis* protein-protein interactome network and the proportion covered by AI-1$_{MAIN}$, its coverage, we calculated the screening completeness, the percentage of all possible *Arabidopsis* pairwise protein combinations screened in space 1 (~10%) (fig. S2), and the overall sensitivity (16%), a parameter that combines both the assay sensitivity of our Y2H version (Fig. 1A and table S5) and the sampling sensitivity of our screens (fig. S5 and table S6) (*3*, *6*, *7*, *9*). Because AI-1$_{MAIN}$ contains 5664 interactions, we estimate that the complete *Arabidopsis* biophysical binary protein-protein interactome, excluding isoforms, is 299,000 ± 79,000 binary interactions (mean ± SD) (*3*), of which AI-1$_{MAIN}$ represents ~2%. Although the *Arabidopsis* interactome is estimated to be larger than those of yeast, worm, or human (*6*, *7*, *9*), the number of interactions per possible protein pairs is similar in all four species (5 to 10 per 10,000). The overall topology of AI-1$_{MAIN}$ is qualitatively similar to that observed for interactome maps of these other species (fig. S6) (*6*, *7*, *9*, *10*). All global network analyses were performed with AI-1MAIN, whereas local analyses were derived from a slightly larger data set, AI-1, obtained by combining AI-1MAIN pairs with interactions identified in repeated screens performed to estimate sampling sensitivity (fig. S2 and tables S4, S6, and S7) (*3*).

Comparing AI-1$_{MAIN}$ to a network of *Arabidopsis* literature-curated interactions. We assembled 4252 literature-curated binary interactions between 2160 *Arabidopsis* proteins (LCI$_{BINARY}$) (fig. S1 and tables S1 and S4) (*3*). The observed overlap with AI-1$_{MAIN}$ lies within the range expected given the AI-1$_{MAIN}$ coverage (Fig. 1B) (*3*). With similar numbers of proteins (nodes)

A



B

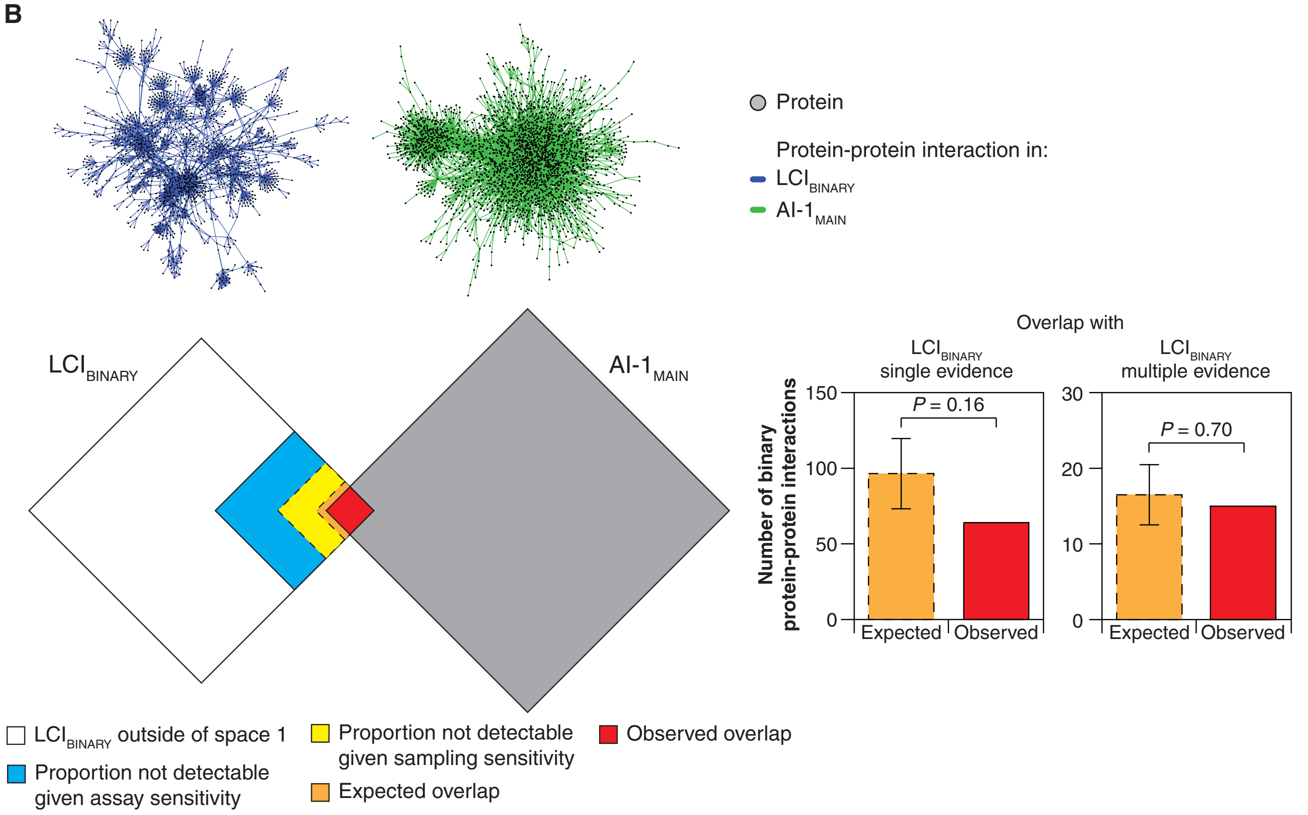


Fig. 1. Quality of AI-1$_{MAIN}$. (**A**) Fraction of PRS, RRS, or AI-1$_{MAIN}$ sample pairs positive in Y2H or in wNAPPA at a scoring threshold of 1.5. Error bars, standard error of the proportion. *P* values, one-sided two-sample *t* tests (*3*). PRS pairs are more often detected than RRS pairs in wNAPPA ($P = 2 \times 10^{-8}$, one-sided two-sample *t* test) and Y2H ($P < 2.2 \times 10^{-16}$, one-sided Fisher's exact test). (**B**) The number of literature-curated interactions recovered reflects AI-1$_{MAIN}$ framework parameters (*6*). (Top) Network representations of LCI$_{BINARY}$ and AI-1$_{MAIN}$. (Bottom left) Data sets are represented by squared Venn diagrams; size is proportional to the number of interactions (*3*). (Bottom right) Observed and expected overlap given sensitivity and completeness of AI-1$_{MAIN}$ with LCIBINARY interactions supported by a single or multiple experimental evidences (*3*). PRS pairs were removed from LCI$_{BINARY}$ multiple evidence for this analysis. Error bars, two SD from the expected counts.

and interactions (edges), AI-1_{MAIN} and LCI_{BINARY} are both small-world networks (fig. S6). However, LCI_{BINARY} shows longer distances between nodes and a higher tendency to form clusters of highly interacting nodes (Fig. 1B and fig. S6). This is likely due to biases inherent to literature-curated data sets, because hypothesis-driven research focuses on a few proteins designated to be important (*5*–*7*, *9*–*11*). AI-1_{MAIN} and LCI_{BINARY} contain similar fractions of plant-specific proteins (19% and 14%, respectively) (fig. S6 and table S8) (*3*), but the presence of several highly connected plant-specific hubs in AI-1_{MAIN} results in twice as many plant-specific interactions (40% and 20%) (fig. S6 and table S9).

Overlap of AI-1 with other biological relationships. To estimate the overall biological relevance of AI-1 interactions, we used statistical correlations with genome-wide functional information available for *Arabidopsis* (*7*, *9*). We observed a significantly higher coexpression correlation for pairs of transcripts encoding interacting proteins than for control pairs (fig. S7) (*3*). Interacting proteins are also enriched in common gene ontology (GO) annotations, particularly those describing specific biological functions and thus assigned to only a few proteins, which we refer to as "precise" annotations (fig. S7) (*3*). This enrichment holds true for GO annotations based strictly on genetic experiments (fig. S7) (*3*). Protein pairs that do not directly interact but share interactors are also enriched in common precise GO annotations (fig. S7) (*3*). Similar to the whole *Arabidopsis* proteome, but in contrast to proteins involved in literature-curated interactions, two-thirds of proteins in AI-1 lack any or precise GO annotations; for these, AI-1 provides starting points for hypothesis development (fig. S7 and tables S8 and S9).

Plant signaling networks in AI-1. Integration of biophysical interactions with orthogonal functional data can uncover novel biological relationships at the scale of individual proteins, pathways, and networks (*1*). We examined ubiquitination enzymes and their substrates, an expanded system in plants relative to other species (*12*). The specific targets of most ubiquitination enzymes remain elusive, and a systems level understanding of ubiquitin signaling is missing. We identified 32 interactions between E3 proteins and potential target proteins shown to be

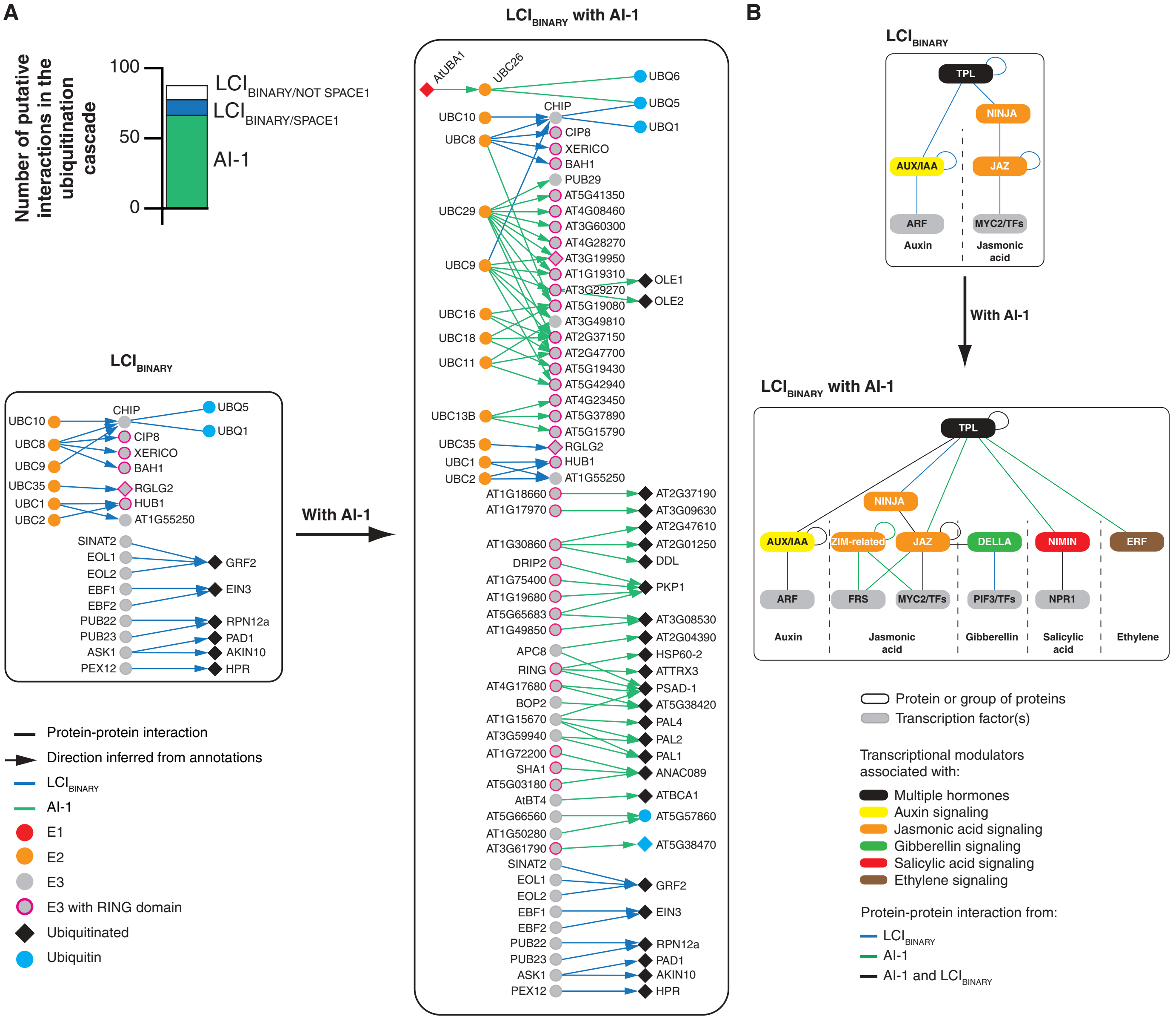

A
Number of putative interactions in the ubiquitination cascade
100
50
0
LCI BINARY/NOT SPACE1
LCI BINARY/SPACE1
AI-1
LCI BINARY
With AI-1
LCI BINARY with AI-1
B
LCI BINARY
TPL
NINJA
AUX/IAA
JAZ
ARF
MYC2/TFs
Auxin
Jasmonic acid
With AI-1
LCI BINARY with AI-1
ZIM-related
DELLA
NIMIN
ERF
FRS
PIF3/TFs
NPR1
Gibberellin
Salicylic acid
Ethylene
Protein-protein interaction
Direction inferred from annotations
LCI BINARY
AI-1
E1
E2
E3
E3 with RING domain
Ubiquitinated
Ubiquitin
Protein or group of proteins
Transcription factor(s)
Transcriptional modulators associated with:
Multiple hormones
Auxin signaling
Jasmonic acid signaling
Gibberellin signaling
Salicylic acid signaling
Ethylene signaling
Protein-protein interaction from:
LCI BINARY
AI-1
AI-1 and LCI BINARY

Fig. 2. Plant signaling networks in AI-1. (**A**) Putative ubiquitination subnetwork extracted from LCI_{BINARY} and AI-1. Bar plot, number of protein-protein interactions between proteins in the ubiquitination cascade in AI-1 and LCI_{BINARY} (outside and within space 1). (**B**) Protein-protein interactions in AI-1 suggest a modular assembly of transcriptional hormone-response regulators and support a global regulatory role for TPL.

ubiquitinated in biochemical experiments (tables S8 and S9) (*3*). Many E3 proteins showed interactions with the same putative target and, conversely, several putative targets interacted with a single common E3 (Fig. 2A) (*3*). Thus, our data support a high combinatorial complexity within the ubiquitination system and, with similar analyses of phosphorylation signaling cascades (fig. S8 and tables S8 and S9) (*3*), provide starting points for analysis of directional information flow through protein-protein interactome networks.

Plant hormones regulate developmental processes and mediate responses to environmental stimuli. In the auxin signaling pathway, auxin/indole-3-acetic acid (AUX/IAA) proteins mediate transcriptional repression of response genes through physical interactions between their ethylene-response-factor–associated amphiphilic repression (EAR) motifs and the co-repressor TOPLESS (TPL) (*13*). Twelve interactions between AUX/IAAs and TPL or TPL-related 3 (TPR3) were observed in AI-1, including six novel ones (fig. S8). Whereas two non-AUX/IAA interactors of TPL have been reported so far (*14*, *15*), there are 21 such interactors in AI-1, of which 15 contain a predicted EAR motif (*16*) ($P < 10^{-24}$, hypergeometric test). TPL interactors include ZIM-domain transcriptional repressors (JAZ5 and JAZ8), regulators of salicylic acid signaling (NIMIN2 and NIMIN3), and a transcriptional regulator of ethylene response (ERF9) (Fig. 2B and fig. S8). AI-1 also reveals direct interactions among repressors, similar to the recently described cross-talk between JAZ proteins and gibberellin-related DELLA proteins (*17*), as well as shared transcription factor targets of JAZ and jasmonic acid–insensitive ZIM-related family members (Fig. 2B and fig. S8). These observations suggest that transcriptional co-repressors and adaptors assemble in a modular way to integrate simultaneous inputs from several hormone pathways and that TPL plays a central role in this process.

Communities in AI-1$_{\text{MAIN}}$. In many networks, communities can be identified as densely interconnected components that function together (*18*). We applied an edge-clustering approach (*19*) to identify communities in AI-1$_{\text{MAIN}}$ and investigated their biological relevance. We identified 26 communities containing more than five proteins in AI-1$_{\text{MAIN}}$ (Fig. 3 and fig. S9) (*3*). About 25% of AI-1$_{\text{MAIN}}$ proteins (661 of 2661) could be assigned to one community, whereas ~1% (23 of 2661) belong to more than one community. We found that ~90% of these communities are enriched in at least one GO annotation (Fig. 3 and table S10) (*3*), whereas negative control networks randomized by degree-preserving edge shuffling showed fewer communities and little GO annotation enrichment ($P < 0.01$, empirical P value) (Fig. 3). Detailed inspection of AI-1$_{\text{MAIN}}$ communities (figs. S10 to S35) both recapitulated available biological information and suggested new hypotheses. For example, the brassinosteroid signaling/phosphoprotein-binding community contains several 14-3-3 proteins known to regulate brassinosteroid signaling (fig. S10). Consistent with the tendency of 14-3-3 proteins to interact with phosphorylated partners (*20*), this community is enriched in experimentally identified phosphoproteins ($P = 0.005$, Fisher's exact test). The interactions between the 14-3-3 proteins and the abscisic acid–responsive element binding transcription factor AREB3 are corroborated by previous findings in barley (*21*) and suggest that plant 14-3-3 proteins mediate multiple hormone signaling pathways.

Several communities, such as transcription/gene expression and nucleosome assembly, share proteins indicating linked biological processes (fig. S36). Particularly striking is the large transmembrane transport community sharing 13 proteins with the vesicle trafficking community

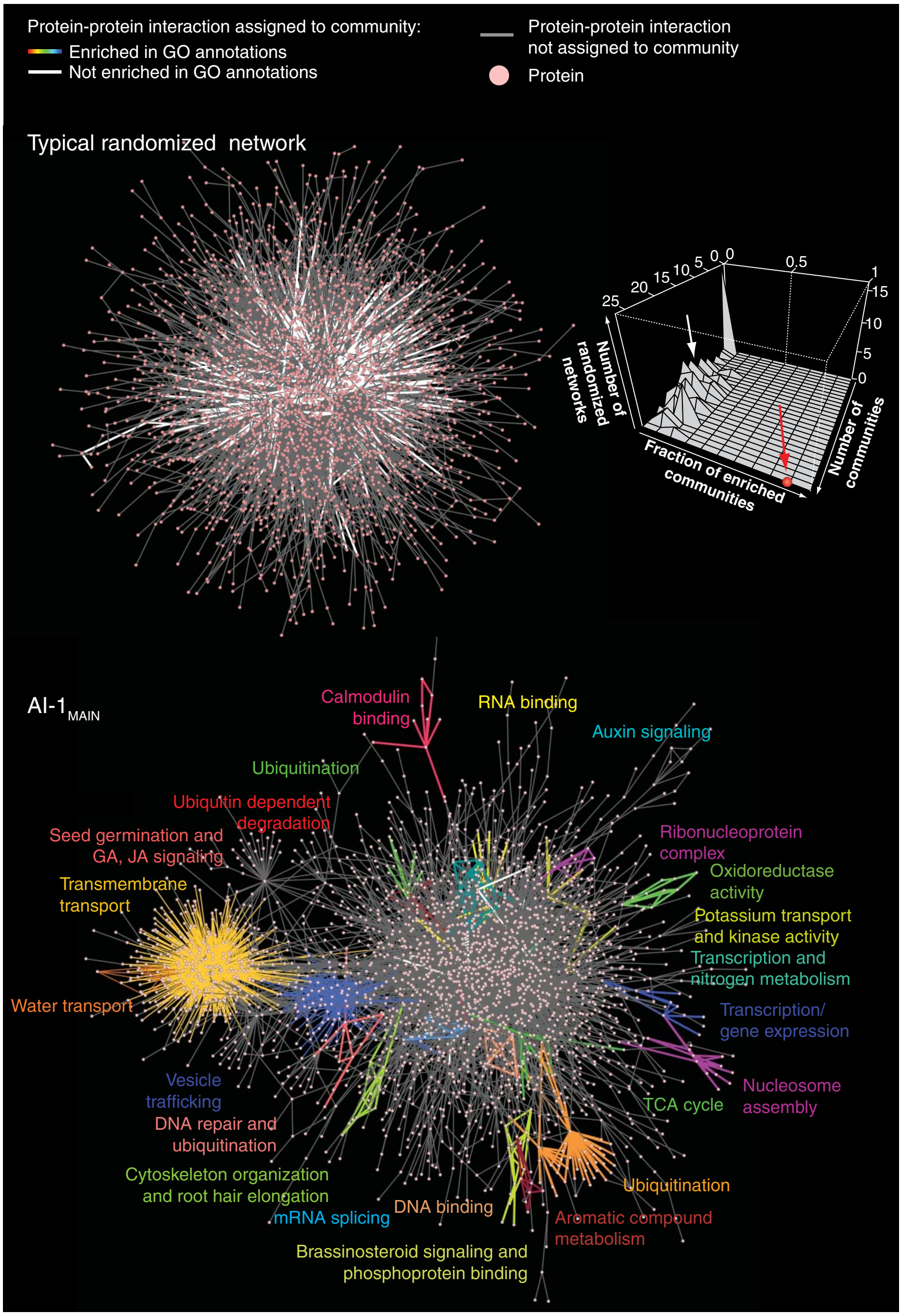

Protein-protein interaction assigned to community:
Enriched in GO annotations
Not enriched in GO annotations
Protein-protein interaction not assigned to community
Protein
Typical randomized network
25
20
15
10
5
0
0
0.5
1
15
10
5
0
Number of randomized networks
Fraction of enriched communities
Number of communities
AI-1$_{\text{MAIN}}$
Calmodulin binding
RNA binding
Auxin signaling
Ubiquitination
Ubiquitin dependent degradation
Seed germination and GA, JA signaling
Transmembrane transport
Ribonucleoprotein complex
Oxidoreductase activity
Potassium transport and kinase activity
Transcription and nitrogen metabolism
Water transport
Transcription/ gene expression
Vesicle trafficking
Nucleosome assembly
TCA cycle
DNA repair and ubiquitination
Cytoskeleton organization and root hair elongation
Ubiquitination
DNA binding
mRNA splicing
Aromatic compound metabolism
Brassinosteroid signaling and phosphoprotein binding

Fig. 3. Communities in AI-1$_{\text{MAIN}}$ (bottom) and in a typical randomized network (top left) (fig. S9). Only the largest connected component of each network is shown. Colored regions indicate communities enriched in GO annotations summarized by the indicated terms (table S10). (Upper right) Distribution of randomized networks as a function of the total number and number of GO annotation enriched communities they contain. White arrow, position of the shown randomized network; red dot and arrow, position of AI-1$_{\text{MAIN}}$. GA, gibberellic acid; JA, jasmonic acid; TCA, tricarboxylic acid.

and six with the water transport community (fig. S36). These shared proteins are bridged by four well-connected proteins within the transmembrane transport community, including two membrane-tethered NAC-type transcription factors, ANAC089 and NTL9 (fig. S36). Transcription factors in this plant-specific protein family are activated by release from the cellular membrane by endopeptidase- or ubiquitin-mediated cleavage (*22*). Interactions corresponding to both mechanisms are found in the transmembrane transport community (fig. S37).

Four distinct communities correspond to ubiquitination. The largest is predominantly composed of interactions between 36 F-box proteins and two Skp proteins, known to form degradative SCF (Skp1, Cullin, F-box) ubiquitin ligase complexes (fig. S27). Two others are composed of shared E2 ubiquitin conjugating enzymes and distinct RING-finger family E3 ligases (figs. S12 and S16). The ubiquitination and DNA repair community includes the UBC13 and MMS2/UEV E2 ubiquitin conjugating enzymes, which participate in nonproteolytic polyubiquitination (fig. S13) (*23*). Distinct types of ubiquitin-related processes were thus identified in AI-1.

Our analyses support the relevance of communities identified in AI-1_{MAIN}, and we anticipate that, with increasing coverage, interactome network maps will improve understanding of the systems-level organization of plants.

Evidence for network evolution. Whether or not natural selection shapes the evolution of interactome networks remains unclear. Gene duplication, a major driving force of evolutionary novelty, has been studied in yeast, providing a framework for understanding subsequent protein-protein interaction rewiring (Fig. 4A) (*24*). However, the difficulty in dating ancient gene duplication events and the low coverage of available protein-protein interaction data sets limit the interpretation of these studies (*3*, *24*–*27*). The high fraction of duplicated genes in the *Arabidopsis* genome compared with nonplant species, combined with the relatively large size of AI-1_{MAIN}, provides interactome data for 1882 paralogous pairs (fig. S38). These pairs span a wide range of apparent interaction rewiring, as measured by the fraction of shared interactors for each pair (fig. S38).

To verify that the apparent interaction rewiring in AI-1_{MAIN} reflects functional divergence, we focused on paralogous pairs classified as having no, low, or high functional divergence on the basis of morphological consequences observed in functionally null mutants of single or pairs of paralogous genes (*28*). For the 17 pairs in AI-1_{MAIN} for which comparative phenotypic data are available, the fraction of shared interactors accurately predicted this functional divergence classification (Fig. 4B).

To study the dynamics of interaction rewiring, we dated gene duplication events using a comparative genomics approach that brackets these events on the basis of multitaxonomic phylogenetic trees (*3*). This allowed us to divide AI-1_{MAIN} paralogous pairs into four time-since-duplication age groups covering up to ~700 million years (fig. S39). To account for the illusion of divergence induced by low experimental coverage, we empirically determined the average fraction of common interactors detected for a set of proteins screened twice, as performed for AI-1_{MAIN} (fig. S40) (*3*). We used this expected upper bound to calibrate the fraction of observed shared interactors between paralogous proteins, assuming that duplicates are identical at the time of duplication (Fig. 4C) (*3*). Our observations

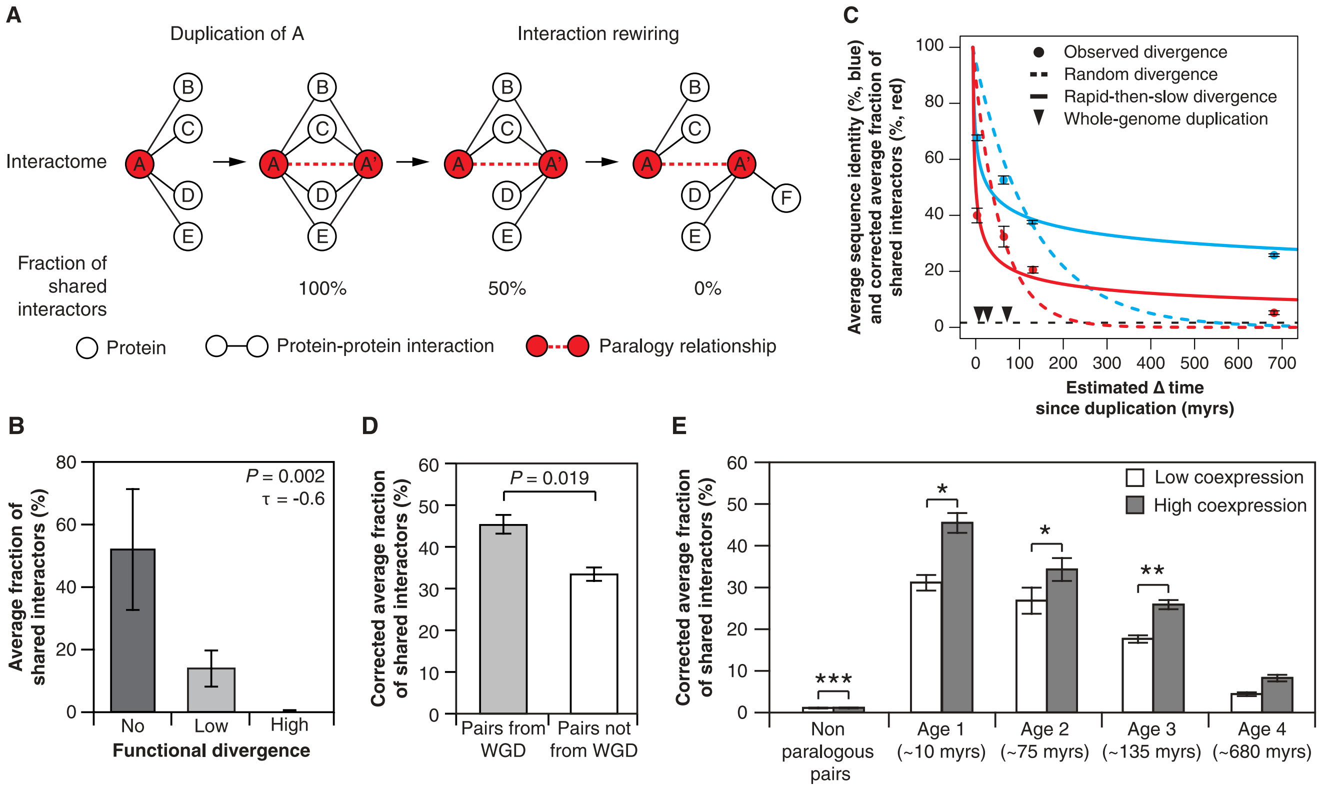


Fig. 4. Evidence for network evolution in AI-1_{MAIN}. (**A**) Interaction rewiring over time, according to the duplication-divergence model (*24*). (**B**) Average fraction of interactors shared between pairs of paralogous proteins with no ($n = 4$), low ($n = 10$), and high ($n = 3$) functional divergence (*28*). Error bars, mean ± SEM. *P* value, one-sided Kendall ranking correlation test (τ, association) (*3*). (**C**) Average fraction of shared interactors, corrected for low experimental coverage (*3*), and average protein sequence identity between pairs of paralogous proteins as a function of the estimated time elapsed since duplication. Error bars, mean ± SEM (*3*). Dashed black line, corrected average fraction of shared interactors of nonparalogous pairs; myrs, million years. (**D**) Corrected average fraction of shared interactors (*3*) for pairs of paralogous proteins originating from polyploidy events ($n = 109$), as compared with other paralogous protein pairs of similar age ($n = 147$). Error bars, mean ± SEM (*3*). *P* values, Mann-Whitney *U* test. (**E**) Corrected average fraction of shared interactors (*3*) for pairs of paralogous proteins encoded by gene pairs with high or low coexpression correlation (top and bottom tertile, respectively) as a function of phylogeny-based age group. Error bars, mean ± SEM (*3*). *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

are not driven by the existence of certain large protein families in $\text{AI-1}_{\text{MAIN}}$ (fig. S41). As reported for yeast (*24*, *26*, *27*), the average fraction of common interactors decreases over evolutionary time, showing substantial and rapid divergence, even after correcting for the coverage of $\text{AI-1}_{\text{MAIN}}$. Yet, in *Arabidopsis*, paralogous pairs that have been diverging for ~700 million years still share more interactors than random protein pairs ($P < 2.2 \times 10^{-16}$, Mann-Whitney *U*-test), indicating that the long-term fate of paralogous proteins is not necessarily a complete divergence of their interaction profiles.

The proportion of shared interactors does not decay exponentially with time-since-duplication, as expected when assuming neutral evolution (*3*, *29*, *30*), that is, random interaction rewiring with no impact on fitness (*31*). Instead, the rate of rewiring appears to be "rapid-then-slow," as suggested by a better fit to a power-law decay (Fig. 4C and fig. S42) (*3*). This trend mirrors that of protein sequence divergence for these paralogous pairs (Fig. 4C), which reflects the variation of selective pressure at different times after the duplication event. After an initial transient relaxation leading to rapid protein sequence divergence, selective pressure tightens on retained paralogs and their divergence decelerates (*3*, *25*) (fig. S39). The fact that interactions diverge in a time-dependent manner similar to protein sequences supports the hypothesis that protein-protein interactions drive the evolution of duplicated genes.

To investigate the interplay between duplication mechanism and the fate of duplicates (*32*), we compared duplicates originating from whole-genome duplications (WGDs) to those from other types of gene duplications. In our most recent age group containing paralogs specific to the *Arabidopsis* genus, 109 paralogous pairs arose during the two most recent WGDs in the *Arabidopsis* lineage (α and β WGDs) (*3*, *33*). As previously observed for yeast (*34*), these pairs share more interactors than other paralogous pairs in the same age group (Fig. 4D and fig. S43), but this effect could simply reflect the younger age of WGD pairs as revealed by more precise time estimates (fig. S43). Although gene dosage balance has been proposed to determine loss or retention of duplicates after WGDs (*33*), the observed extensive rewiring reinforces previous observations pointing to functional divergence as a major feature of the long-term evolution of polyploid plants (*35*).

Expression profile divergence is rapid, nonrandom, and substantial in *Arabidopsis* (*36*, *37*) (fig. S44), yet appears to play a limited role in the functional divergence of paralogs (*28*). We tested whether the evolutionary forces acting on expression profiles and protein interaction divergence are complementary or correlated. For each duplication age group, the most coexpressed paralogous pairs tend to share more interactors than the least coexpressed ones (Fig. 4E). This suggests that selective pressures driving functional divergence concurrently act on both aspects of protein function.

With >65% sequence identity and strongly correlated expression profiles, the most recent paralogous pairs share less than half of their interactors (41%) (Fig. 4C and figs. S44 and S45). This contrast is consistent with the common understanding that protein-protein interactions are only one of many constraints limiting sequence changes during evolution, allowing for small sequence changes to induce fate-determining network rewiring (*38*, *39*). One example of interaction rewiring despite sequence conservation is observed in the actin family. Each actin protein pair shares >90% sequence identity, yet collectively the actin family exhibits time-dependent interaction rewiring (fig. S45).

Modeling interaction rewiring with nonconstant rates should provide insight into the evolution of interactome networks and their topology (*40*). Whether this rewiring is merely a consequence of sequence divergence or is a primary driver remains an open question. Together with observations of fast rewiring of other types of biological networks (*41*, *42*), our data invite speculation that edge-specific rewiring is faster than node evolution in biological networks.

Conclusion. Our empirically determined high-quality protein-protein interaction map for a plant interactome network should not only hasten the functional characterization of unknown proteins, including those with potential biotechnological utility, but also enable systems level investigations of genotype-to-phenotype relationships in the plant kingdom. One example is how AI-1 illuminates mechanisms and strategies by which plants cope with pathogenic challenges (*43*).

The paradigms established here are compatible with models in which the interactome network constrains and shapes sequence evolution. Studying sequence variation, conservation, mutation, and evolution rate has shed light on how natural selection drives evolution. Explorations of interaction variation will similarly broaden the understanding of network evolution, whether in the context of duplication or trans-kingdom comparative interactomics.

References and Notes

1. M. Vidal, M. E. Cusick, A. L. Barabási, *Cell* **144**, 986 (2011).
2. K. Yamada *et al.*, *Science* **302**, 842 (2003).
3. See Supporting Online Material for a detailed description.
4. M. Dreze *et al.*, *Methods Enzymol.* **470**, 281 (2010).
5. M. E. Cusick *et al.*, *Nat. Methods* **6**, 39 (2009).
6. K. Venkatesan *et al.*, *Nat. Methods* **6**, 83 (2009).
7. N. Simonis *et al.*, *Nat. Methods* **6**, 47 (2009).
8. P. Braun *et al.*, *Nat. Methods* **6**, 91 (2009).
9. H. Yu *et al.*, *Science* **322**, 104 (2008).
10. J. F. Rual *et al.*, *Nature* **437**, 1173 (2005).
11. A. M. Edwards *et al.*, *Nature* **470**, 163 (2011).
12. E. Mazzucotelli *et al.*, *Curr. Genomics* **7**, 509 (2006).
13. N. T. Krogan, J. A. Long, *Curr. Opin. Plant Biol.* **12**, 628 (2009).
14. M. Kieffer *et al.*, *Plant Cell* **18**, 560 (2006).
15. L. Pauwels *et al.*, *Nature* **464**, 788 (2010).
16. S. Kagale, M. G. Links, K. Rozwadowski, *Plant Physiol.* **152**, 1109 (2010).
17. X. Hou, L. Y. Lee, K. Xia, Y. Yan, H. Yu, *Dev. Cell* **19**, 884 (2010).
18. S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
19. Y. Y. Ahn, J. P. Bagrow, S. Lehmann, *Nature* **466**, 761 (2010).
20. D. Bridges, G. B. Moorhead, *Sci. STKE* **2005**, re10 (2005).
21. P. J. Schoonheim *et al.*, *Plant J.* **49**, 289 (2007).
22. P. J. Seo, S. G. Kim, C. M. Park, *Trends Plant Sci.* **13**, 550 (2008).
23. R. Wen *et al.*, *Plant Cell* **20**, 213 (2008).
24. A. Wagner, *Mol. Biol. Evol.* **18**, 1283 (2001).
25. M. Lynch, J. S. Conery, *Science* **290**, 1151 (2000).
26. A. Wagner, *Proc. Biol. Sci.* **270**, 457 (2003).
27. S. Maslov, K. Sneppen, K. A. Eriksen, K. K. Yan, *BMC Evol. Biol.* **4**, 9 (2004).
28. K. Hanada, T. Kuromori, F. Myouga, T. Toyoda, K. Shinozaki, *PLoS Genet.* **5**, e1000781 (2009).
29. R. Pastor-Satorras, E. Smith, R. V. Solé, *J. Theor. Biol.* **222**, 199 (2003).
30. A. Vázquez, A. Flammini, A. Maritan, A. Vespignani, *Complexus* **1**, 38 (2003).
31. E. D. Levy, C. R. Landry, S. W. Michnick, *Sci. Signal.* **2**, pe11 (2009).
32. H. Innan, F. Kondrashov, *Nat. Rev. Genet.* **11**, 97 (2010).
33. M. Freeling, *Annu. Rev. Plant Biol.* **60**, 433 (2009).
34. Y. Guan, M. J. Dunham, O. G. Troyanskaya, *Genetics* **175**, 933 (2007).
35. G. Blanc, K. H. Wolfe, *Plant Cell* **16**, 1679 (2004).
36. T. Casneuf, S. De Bodt, J. Raes, S. Maere, Y. Van de Peer, *Genome Biol.* **7**, R13 (2006).
37. E. W. Ganko, B. C. Meyers, T. J. Vision, *Mol. Biol. Evol.* **24**, 2298 (2007).
38. C. Shou *et al.*, *PLOS Comput. Biol.* **7**, e1001050 (2011).
39. M. Dreze *et al.*, *Nat. Methods* **6**, 843 (2009).
40. A. L. Barabási, Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).
41. G. D. Amoutzias *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2967 (2010).
42. A. E. Mayo, Y. Setty, S. Shavit, A. Zaslaver, U. Alon, *PLoS Biol.* **4**, e45 (2006).
43. M. S. Mukhtar *et al.*, *Science* **333**, 596 (2011).

Acknowledgments: We thank P. Benfey, H. Yu, M. Nordborg, P. Ronald, M Snyder, and R. Wing as well as members of the Dana-Farber Cancer Institute Center for Cancer Systems Biology, for helpful discussions. This work was supported by the following grants: NSF 0703905 to M.V., J.R.E., and D.E.H.; National Human Genome Research Institute R01HG001715 to M.V., D.E.H., and F.P.R.; NSF 0520253 and NSF 0313578 to J.R.E.; Canada Excellence Research Chairs Program and Canadian Institute for Advanced Research Fellowship to F.P.R.; James S. McDonnell Foundation 220020084 to A.-L.B; Sixth Framework Programme LSHG-CT-2006-037704 (AGRON-OMICS) to C.L.; National Institute of General Medical Sciences R01GM066025 to J.L.D.; U.S. Department of Agriculture, Agricultural Research Service 1907-21000-030 to D.W.; NIH National Research Service Award fellowships F32HG004098 to M.T. and F32HG004830 to R.J.S.; Biotechnology and Biological Sciences Research Council grant F005806 to Jim Beynon in support of J.M.; and NSF 0703908 to D.W. in support of J.S. and W.S. M.V. is a Chercheur Qualifié Honoraire from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation, Belgium). Data reported here are available at the Web site http://interactome.dfci.harvard.edu/A_thaliana

***Arabidopsis* Interactome Mapping Consortium**

Authorship of this paper should be cited as "*Arabidopsis* Interactome Mapping Consortium." Participants are arranged by working group, then listed in alphabetical order, except for chairs, co-chairs, and project leaders when indicated. Matija Dreze, Anne-Ruxandra Carvunis, Benoit Charloteaux, Mary Galli, Samuel J. Pevzner, and Murat Tasan contributed equally to this work and should be considered co-first authors.

Steering group: Pascal Braun[1,2]† (chair), Anne-Ruxandra Carvunis,[1,2,3] Benoit Charloteaux,[1,2,4] Matija Dreze,[1,2,5] Joseph R. Ecker,[6,7]† David E. Hill,[1,2]† Frederick P. Roth,[1,8]‡ Marc Vidal[1,2]†. **ORFeome group:** Mary Galli[6] (project leader), Padmavathi Balumuri,[9] Vanessa Bautista,[6] Jonathan D. Chesnut,[9] Rosa Cheuk Kim,[6]§ Chris de los Reyes,[6] Patrick Gilles,[9]|| Christopher J. Kim,[6] Uday Matrubutham,[9] Jyotika Mirchandani,[9] Eric Olivares,[9]¶

Suswapna Patnaik,[9] Rosa Quan,[6] Gopalakrishna Ramaswamy,[9]# Paul Shinn,[6] Geetha M. Swamilingiah,[9] Stacy Wu,[6] Joseph R. Ecker[6,7]† (chair).

Interactome data acquisition group: Matija Dreze[1,2,5] (project leader), Danielle Byrdsong,[1,2] Amélie Dricot,[1,2] Melissa Duarte,[1,2] Fana Gebreab,[1,2] Bryan J. Gutierrez,[1,2] Andrew MacWilliams,[1,2] Dario Monachello,[12]** M. Shahid Mukhtar,[11]†† Matthew M. Poulin,[1,2] Patrick Reichert,[1,2] Viviana Romero,[1,2] Stanley Tam,[1,2] Selma Waaijers,[1,2]‡‡ Evan M. Weiner,[1,2] Marc Vidal[1,2]† (co-chair), David E. Hill[1,2]† (co-chair), Pascal Braun[1,2]† (chair).

wNAPPA interactome validation group: Mary Galli[6] (project leader), Anne-Ruxandra Carvunis,[1,2,3] Michael E. Cusick,[1,2] Matija Dreze,[1,2,5] Viviana Romero,[1,2] Frederick P. Roth,[1,8]‡ Murat Tasan,[8] Junshi Yazaki,[7] Pascal Braun[1,2]† (co-chair), Joseph R. Ecker[6,7]† (chair).

Bioinformatics and analysis group: Anne-Ruxandra Carvunis[1,2,3] (project leader), Yong-Yeol Ahn,[1,10] Albert-László Barabási,[1,10] Benoit Charloteaux,[1,2,4] Huaming Chen,[6] Michael E. Cusick,[1,2] Jeffery L. Dangl,[11] Matija Dreze,[1,2,5] Joseph R. Ecker,[6,7]† Changyu Fan,[1,2] Lantian Gai,[6] Mary Galli,[6] Gourab Ghoshal,[1,10] Tong Hao,[1,2] David E. Hill,[1,2]† Claire Lurin,[12] Tijana Milenkovic,[13] Jonathan Moore,[14] M. Shahid Mukhtar,[11]†† Samuel J. Pevzner,[1,2,15,16] Natasa Przulj,[17] Sabrina Rabello,[1,10] Edward A. Rietman,[1,2]§§ Thomas Rolland,[1,2] Frederick P. Roth,[1,8]‡ Balaji Santhanam,[1,2] Robert J. Schmitz,[7] William Spooner,[18,19] Joshua Stein,[18] Murat Tasan,[8] Jean Vandenhaute,[5] Doreen Ware,[18,20] Pascal Braun[1,2]† (co-chair), Marc Vidal[1,2]† (chair).

Writing group: Pascal Braun[1,2]† (chair), Anne-Ruxandra Carvunis,[1,2,3] Benoit Charloteaux,[1,2,4] Matija Dreze,[1,2,5] Mary Galli,[6] Marc Vidal[1,2]† (co-chair).

[1]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [3]Computational and Mathematical Biology Group, Techniques de l'Ingénierie Médicale et de la Complexité–Informatique, Mathématiques et Applications de Grenoble, CNRS UMR5525 and Université de Grenoble, Faculté de Médecine, 38706 La Tronche Cedex, France. [4]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Wallonia-Brussels Federation, Belgium. [5]Unité de Recherche en Biologie Moléculaire, Facultés Universitaires Notre-Dame de la Paix, 5000 Namur, Wallonia-Brussels Federation, Belgium. [6]Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [7]Plant Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [8]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [9]Life Technologies, Carlsbad, CA 92008, USA. [10]Center for Complex Network Research (CCNR), Department of Physics, Northeastern University, Boston, MA 02115, USA. [11]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [12]Unité de Recherche en Génomique Végétale, Unités Mixtes de Recherche Institut Nationale de la Recherche Agronomique/Université Evry-Val d'Essonne ERL Centre National de la Recherche Scientifique 91057, Evry Cedex, France. [13]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA. [14]Warwick Systems Biology Centre, Coventry House, University of Warwick, Coventry, CV4 7AL, UK. [15]Biomedical Engineering Department, Boston University, Boston, MA 02215, USA. [16]Boston University School of Medicine, Boston, MA 02118, USA. [17]Department of Computing, Imperial College London, London SW7 2AZ, UK. [18]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [19]Eagle Genomics Ltd., Babraham Research Campus, Cambridge, CB4 1JD, UK. [20]United States Department of Agriculture, Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY 14853, USA.

‡Present address: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S3E1, Canada and Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto, Ontario M5G1X5, Canada.
§Present address: Foley & Lardner LLP, 3579 Valley Centre Drive, Suite 300, San Diego, CA 92130, USA.
‖Deceased.
¶Present address: Pacific Biosciences, 940 Hamilton Drive, Menlo Park, CA 94025, USA.
#Present address: Thermo Fisher Scientific, BioSciences Division, Bangalore-560011, India.
**Present address: Centre de Génétique Moléculaire du C.N.R.S., 1 Avenue de la Terrasse, 91190 Gif-sur-Yvette, France.
††Present address: Department of Biology, University of Alabama at Birmingham, Birmingham, AL 35294, USA.
‡‡Present address: University of Utrecht, 3508 TC Utrecht, The Netherlands.
§§Present address: Center of Cancer Systems Biology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, MA 02135, USA.

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6042/601/DC1
SOM Text
Figs. S1 to S47
Tables S1 to S12
References

7 February 2011; accepted 10 June 2011
10.1126/science.1203877

Friction Anisotropy–Driven Domain Imaging on Exfoliated Monolayer Graphene

Jin Sik Choi,[1] Jin-Soo Kim,[1] Ik-Su Byun,[1] Duk Hyun Lee,[1] Mi Jung Lee,[1] Bae Ho Park,[1]* Changgu Lee,[2] Duhee Yoon,[3] Hyeonsik Cheong,[3] Ki Ho Lee,[4] Young-Woo Son,[4] Jeong Young Park,[5]* Miquel Salmeron[6]

Graphene produced by exfoliation has not been able to provide an ideal graphene with performance comparable to that predicted by theory, and structural and/or electronic defects have been proposed as one cause of reduced performance. We report the observation of domains on exfoliated monolayer graphene that differ by their friction characteristics, as measured by friction force microscopy. Angle-dependent scanning revealed friction anisotropy with a periodicity of 180° on each friction domain. The friction anisotropy decreased as the applied load increased. We propose that the domains arise from ripple distortions that give rise to anisotropic friction in each domain as a result of the anisotropic puckering of the graphene.

The mechanical exfoliation method to transfer a monolayer of graphene to a substrate is thought to be a facile method to obtain a single crystalline graphene (*1*). Mechanical exfoliation, however, may induce strain on the graphene layer during deposition on a substrate and can create wrinkled films and other defects, because the interaction with the substrate might introduce uneven compressive and tensile stresses that are nonuniformly distributed across the film. Structural defects such as atomic defects (*2*), wrinkles or ripples (*3*–*5*), and microscopic corrugation (*6*) have already been reported on supported graphene. These defects tend to lower the electrical performance of graphene devices because they break translational or rotational symmetry. In addition, the boundaries of microscale domains also break the symmetry, as reported for graphene grown by chemical vapor deposition (*7*). However, no experimental observations of microscale domains on mechanically exfoliated monolayer graphene have been reported to date.

Atomic force microscopy (AFM) can be used to study the mechanical properties of surfaces because it provides local information about hardness, deformation, slipperiness, and chemical state. Friction force microscopy (FFM) has been used to investigate elastic deformation, atomic structure, dislocation, and defects (*8*, *9*). Recent studies show that friction depends on the number of graphene layers as well as the nature of the graphene-substrate bond (*10*–*12*). Puckering induced by AFM tip scanning has been proposed as the origin of the thickness effects on friction (*10*). Here, we show the existence of domains on exfoliated monolayers of graphene deposited on silicon oxide that are distinguished by their different friction characteristics when an AFM tip slides over them. These domains cannot be observed in AFM topographic images, optical microscopy, or micro-Raman spectroscopy.

---

[1]Division of Quantum Phases and Devices, Department of Physics, Konkuk University, Seoul 143-701, Korea. [2]Department of Mechanical Engineering, Sungkyunkwan University, Suwon 440-746, Korea. [3]Department of Physics, Sogang University, Seoul 121-742, Korea. [4]Korea Institute for Advanced Study, Seoul 130-722, Korea. [5]Graduate School of Energy, Environment, Water, and Sustainability, NanoCentury KI, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea. [6]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: baehpark@konkuk.ac.kr (B.H.P.); jeongypark@kaist.ac.kr (J.Y.P.)