Evidence for Network Evolution in an *Arabidopsis* Interactome Map

Arabidopsis Interactome Mapping Consortium[*]

Abstract

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 ~6,200 highly reliable interactions between ~2,700 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, tables S1, S2), depriving us of an understanding of how genotype-to-phenotype relationships are mediated at the systems level (1).

A high-quality binary protein-protein interactome map for *Arabidopsis*

To generate a map of the *Arabidopsis* interactome network, we used a collection of ~8,000 open reading frames representing ~30% of its predicted protein-coding genes (table S3) (2, 3). We tested all pair-wise 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 dataset of 5,664 binary interactions between 2,661 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 re-curated (5) *Arabidopsis* protein-protein interactions and a random reference set (RRS) of 146 random protein pairs (fig. S3; 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; table S5) (3, 8).

[*]List of participants and affiliations appear at the end of the manuscript.

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* pair-wise protein combinations screened in space 1 (~10%) (fig. S2), and the *overall sensitivity,* a parameter that combines both the assay sensitivity of our Y2H version (Fig. 1A) and the sampling sensitivity of our screens (~16%) (fig. S5; table S5) (3, 6, 7, 9). Since AI-1$_{MAIN}$ contains 5,664 interactions, we estimate that the complete *Arabidopsis* biophysical binary protein-protein interactome, excluding isoforms, is 299,000 ± 79,000 binary interactions (mean ± standard deviation) (3), of which AI-1$_{MAIN}$ represents ~2%. While 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 (5–10 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). While all global network analyses were performed with AI-1$_{MAIN}$, local network analyses were done with AI-1 (http://interactome.dfci.harvard.edu/A_thaliana/index.php?page=2010anm_download and http://interactome/A_thaliana/index.php?page=display; table S4), a dataset combining AI-1$_{MAIN}$ and interactions identified in repeated screens on the subspace indicated in fig. S2, performed to estimate sampling sensitivity (tables S4, S6, S7) (3).

Comparing AI-1$_{MAIN}$ to a network of *Arabidopsis* literature-curated interactions

We assembled 4,252 literature-curated binary interactions between 2,160 *Arabidopsis* proteins (LCI$_{BINARY}$) (fig. S1; tables S1, 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) 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) (fig. S6). This is likely due to biases inherent to literature-curated datasets, as hypothesis-driven research focuses on 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; 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; 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 co-expression 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" (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; tables S8, 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 ubiquitinated in biochemical experiments (tables S8, 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; tables S8, 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 via physical interactions between their ethylene-response-factor-associated amphiphilic repression (EAR) motifs and the co-repressor TOPLESS (TPL) (13). Twelve interactions between AUX/IAA and TPL or TPL-related3 (TPR3) were observed in AI-1, including six novel ones (fig. S8). While 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, JAZ8), regulators of salicylic acid signaling (NIMIN2, NIMIN3), and a transcriptional regulator of ethylene response (ERF9) (Fig. 2B, fig. S8). AI-1 also reveals direct interactions among co-repressors, similar to the recently described crosstalk 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; 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$_{MAIN}$

In many networks, communities can be identified with densely interconnected components that function together (18). We applied an edge clustering approach (19) to identify communities in AI-1$_{MAIN}$ and investigated their biological relevance. We identified 26 communities containing more than five proteins in AI-1$_{MAIN}$ (Fig. 3; fig. S9) (3). Approximately 25% of AI-1$_{MAIN}$ proteins (661/2,661) could be assigned to one community, while ~1% (23/2,661) belong to more than one community. We found that ~90% of these communities are enriched in at least one GO annotation (Fig. 3; table S10) (3), whereas negative control networks randomized by degree-preserving edge shuffling showed fewer communities and little GO annotation enrichment ($P < 0.01$; Fig. 3). Detailed inspection of AI-1$_{MAIN}$ communities (figs. S10–35) 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" 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 mediated

transport" community and six with the "water transport" community (fig. S36). These shared proteins are bridged via 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, S16). The "ubiquitination and DNA repair" community includes the UBC13 and MMS2/UEV E2 ubiquitin conjugating enzymes, which participate in non-proteolytic 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 our understanding of the systems-level molecular 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 to date ancient gene duplication events and the low coverage of available protein-protein interaction datasets limit the interpretation of these studies (3, 24–27). The high fraction of duplicated genes in the *Arabidopsis* genome compared to non-plant species, combined with the relatively large size of AI-1$_{MAIN}$, provides interactome data for 1,882 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 is 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 multi-taxonomic 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 are not driven by the existence of certain large protein families in AI-1$_{MAIN}$ (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 AI-1$_{MAIN}$. Yet, in *Arabidopsis*, paralogous pairs that have been diverging for

~700 million years still share more interactors than random proteins 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), *i.e.* random interaction rewiring, with no impact on fitness (31). Instead, the rate of rewiring appears "rapid-then-slow", as suggested by a better fit to a power-law decay (Fig. 4C; 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; fig. S43), but this effect could simply reflect the younger age of WGD pairs as revealed by more precise time estimates (fig. S43). While gene dosage balance has been proposed to determine loss or retention of duplicates following 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, non-random 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 co-expressed paralogous pairs tend to share more interactors than the least co-expressed 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; figs. S44, 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 non-constant 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 (Mukhtar *et al.*, co-submitted).

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Vidal M, Cusick ME, Barabasi AL. Cell. 2011; 144:986. [PubMed: 21414488]

2. Yamada K, et al. Science. 2003; 302:842. [PubMed: 14593172]

3. See Supplementary Online Material for a detailed description.

4. Dreze M, et al. Methods Enzymol. 2010; 470:281. [PubMed: 20946815]

5. Cusick ME, et al. Nat Methods. 2009; 6:39. [PubMed: 19116613]

6. Venkatesan K, et al. Nat Methods. 2009; 6:83. [PubMed: 19060904]

7. Simonis N, et al. Nat Methods. 2009; 6:47. [PubMed: 19123269]

8. Braun P, et al. Nat Methods. 2009; 6:91. [PubMed: 19060903]

9. Yu H, et al. Science. 2008; 322:104. [PubMed: 18719252]

10. Rual JF, et al. Nature. 2005; 437:1173. [PubMed: 16189514]

11. Edwards AM, et al. Nature. 2010:470.

12. Mazzucotelli E, et al. Curr Genomics. 2006; 7:509. [PubMed: 18369404]

13. Krogan NT, Long JA. Curr Opin Plant Biol. 2009; 12:628. [PubMed: 19700365]

14. Kieffer M, et al. Plant Cell. 2006; 18:560. [PubMed: 16461579]

15. Pauwels L, et al. Nature. 2010; 464:788. [PubMed: 20360743]

16. Kagale S, Links MG, Rozwadowski K. Plant Physiol. 2010; 152:1109. [PubMed: 20097792]

17. Hou X, Lee LY, Xia K, Yan Y, Yu H. Dev Cell. 2010; 19:884. [PubMed: 21145503]

18. Fortunato S. Phys Rep. 2010; 486:75.

19. Ahn YY, Bagrow JP, Lehmann S. Nature. 2010; 466:761. [PubMed: 20562860]

20. Bridges D, Moorhead GB. Sci STKE. 2005; 296:10.

21. Schoonheim PJ, et al. Plant J. 2007; 49:289. [PubMed: 17241451]

22. Seo PJ, Kim SG, Park CM. Trends Plant Sci. 2008; 13:550. [PubMed: 18722803]

23. Wen R, et al. Plant Cell. 2008; 20:213. [PubMed: 18178771]

24. Wagner A. Mol Biol Evol. 2001; 18:1283. [PubMed: 11420367]

25. Lynch M, Conery JS. Science. 2000; 290:1151. [PubMed: 11073452]

26. Wagner A. Proc Biol Sci. 2003; 270:457. [PubMed: 12641899]

27. Maslov S, Sneppen K, Eriksen KA, Yan KK. BMC Evol Biol. 2004; 4:9. [PubMed: 15070432]

28. Hanada K, Kuromori T, Myouga F, Toyoda T, Shinozaki K. PLoS Genet. 2009; 5:e1000781. [PubMed: 20041196]

29. Pastor-Satorras R, Smith E, Sole RV. J Theor Biol. 2003; 222:199. [PubMed: 12727455]

30. Vazquez A, Flammini A, Maritan A, Vespignani A. ComPlexUs. 2003; 1:38.

31. Levy ED, Landry CR, Michnick SW. Sci Signal. 2009; 2:pe11. [PubMed: 19261595]

32. Innan H, Kondrashov F. Nat Rev Genet. 2010; 11:97. [PubMed: 20051986]

33. Freeling M. Annu Rev Plant Biol. 2009; 60:433. [PubMed: 19575588]

34. Guan Y, Dunham MJ, Troyanskaya OG. Genetics. 2007; 175:933. [PubMed: 17151249]

35. Blanc G, Wolfe KH. Plant Cell. 2004; 16:1679. [PubMed: 15208398]

36. Casneuf T, De Bodt S, Raes J, Maere S, Van de Peer Y. Genome Biol. 2006; 7:R13. [PubMed: 16507168]

37. Ganko EW, Meyers BC, Vision TJ. Mol Biol Evol. 2007; 24:2298. [PubMed: 17670808]

38. Shou C, et al. PLoS Comput Biol. 2011; 7:e1001050. [PubMed: 21253555]

39. Dreze M, et al. Nat Methods. 2009; 6:843. [PubMed: 19855391]

40. Barabasi AL, Oltvai ZN. Nat Rev Genet. 2004; 5:101. [PubMed: 14735121]

41. Amoutzias GD, et al. Proc Natl Acad Sci USA. 2010; 107:2967. [PubMed: 20080574]

42. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. PLoS Biol. 2006; 4:e45. [PubMed: 16602820]

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 (AO), except for Chairs, co-Chairs and Project Leaders when indicated.

Correspondence and request for materials should be addressed, to M.V. (marc_vidal@dfci.harvard.edu); J.R.E. (ecker@salk.edu); P.B. (pascal_braun@dfci.harvard.edu); D.E.H. (david_hill@dfci.harvard.edu).

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 (AO): 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).

NAPPA 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 (AO): 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).

[††]Present address: Center of Cancer Systems Biology, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, MA 02135, USA.
[18]Cold Spring Harbor Laboratory (CSHL), 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 (USDA ARS), Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY 14853, USA.
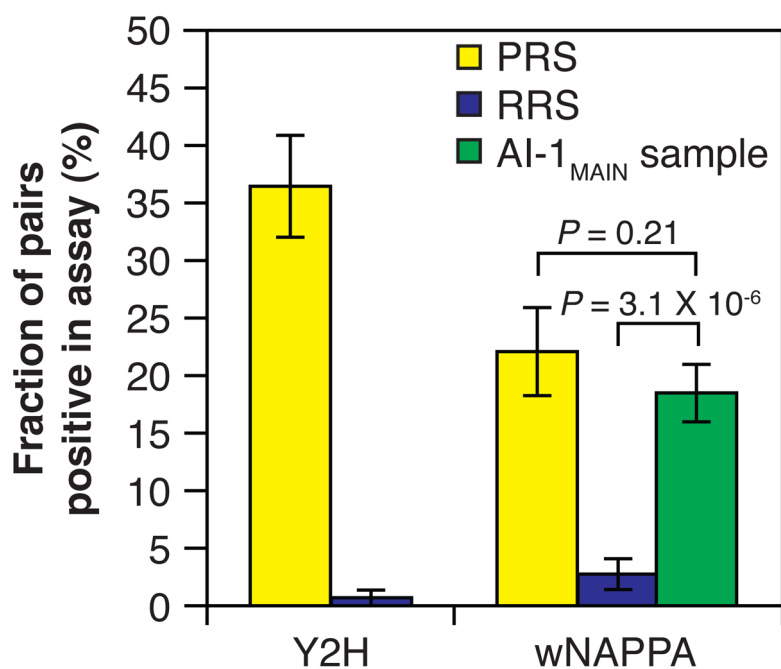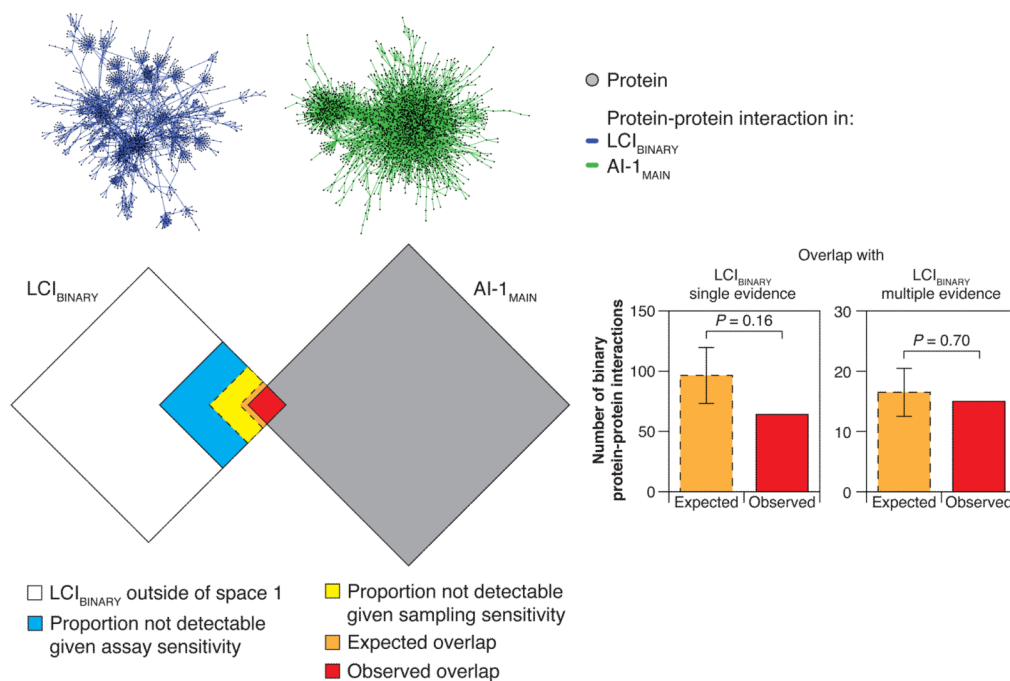
Figure 1A

Figure 1B

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$_{MAIN}$ (see main text and (3)). PRS pairs were removed from LCI$_{BINARY}$ multiple evidence for this analysis. Error bars: two standard deviations from the expected counts.
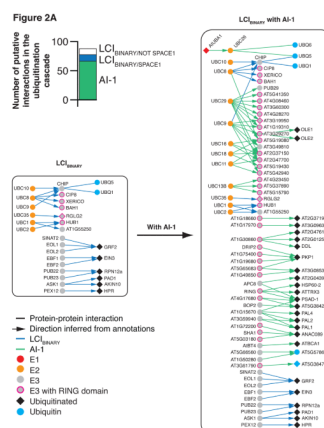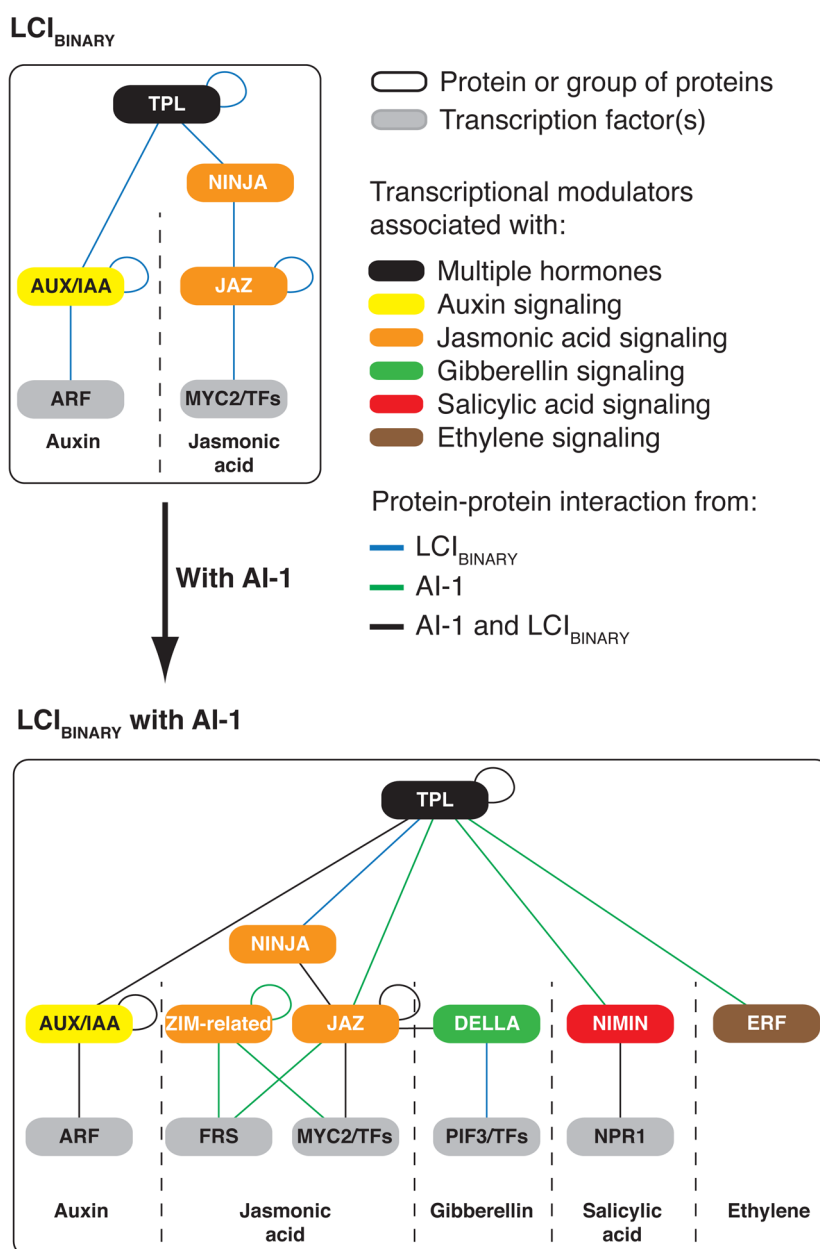
Figure 2A

Figure 2B

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 LCI$_{BINARY}$ and AI-1 (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.
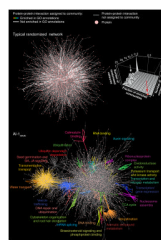
Fig. 3.
Communities in AI-1$_{MAIN}$ (bottom) and in a typical randomized network (top left; fig. S9). Only largest connected component of networks are shown. Colored regions indicate communities enriched in GO annotations summarized by the indicated terms (table S10). Upper right: distribution of randomized networks as a function of the total and GO annotation enriched number of communities they contain; white arrow: position of the shown randomized network, red dot and arrow: position of AI-1$_{MAIN}$. GA: gibberellic acid, JA: jasmonic acid, TCA: tricarboxylic acid.
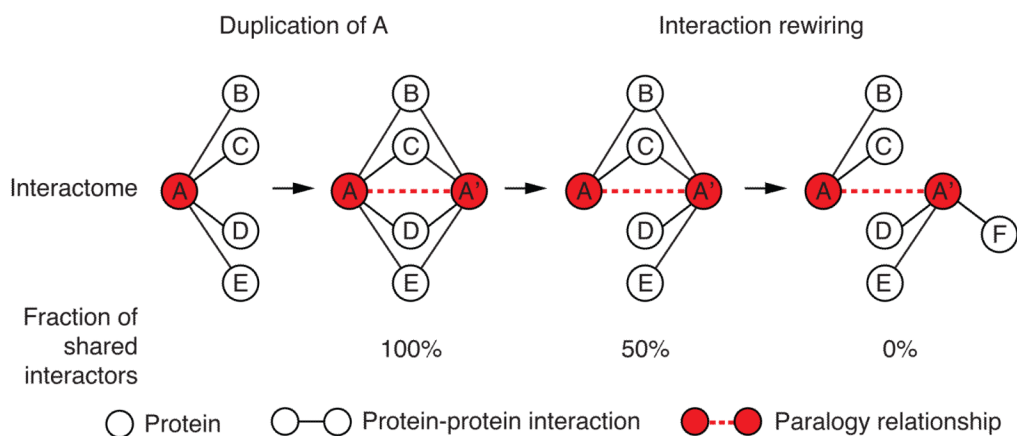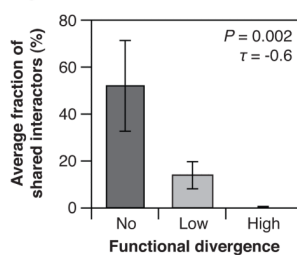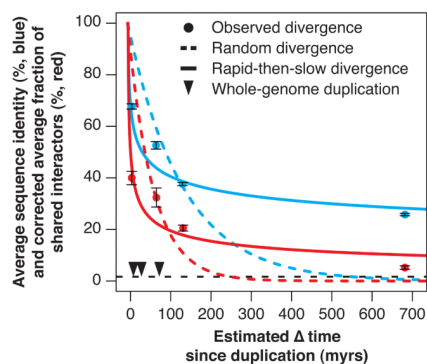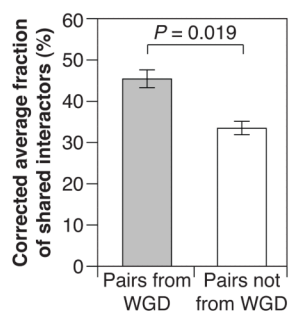
Figure 4A

Duplication of A   Interaction rewiring

Interactome

Fraction of
shared
interactors

100%   50%   0%

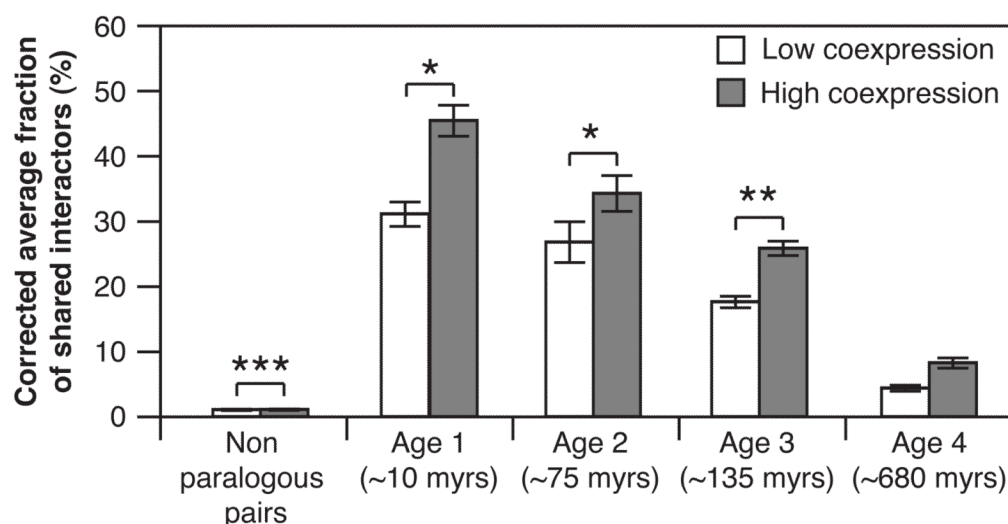○ Protein   ○—○ Protein-protein interaction   ●--● Paralogy relationship

Figure 4B

Figure 4C

Figure 4D

Figure 4E

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: standard error of the mean. *P*-value: one-sided Kendall ranking correlation test ($\tau=$ 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 Δ time-since-duplication. Error bars: standard error of the mean (3). Dashed black line: corrected average fraction of shared interactors of non-paralogous 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 to other paralogous protein pairs of similar age ($n=147$). Error bars: standard error of the mean (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 co-expression correlation (top and bottom tertile, respectively) as a function of phylogeny-based age group. Error bars: standard error of the mean (3). $P < 0.05$ (*), < 0.01 (**), < 0.001 (***).