

Evidence of Interaction Network Evolution by Whole-Genome Duplications: A Case Study in MADS-Box Proteins

Amelie S. Veron,* Kerstin Kaufmann,† and Erich Bornberg-Bauer*

*Division of Bioinformatics, Institute for Evolution and Biodiversity, The Westphalian Wilhelms University of Münster, Münster, Germany; and †Business Unit Bioscience, Plant Research International, Wageningen, The Netherlands

Recent investigations on metazoan transcription factors (TFs) indicate that single-gene duplication events and the gain and loss of protein domains are 2 crucial factors in shaping their protein–protein interaction networks. Plant genomes, on the other hand, have a history of polyploidy and whole-genome duplications (WGDs), and thus, their study helps to understand whether WGDs have also had a significant influence on protein network evolution. Here we investigate the evolution of the interaction network in the well-studied MADS domain MIKC-type proteins, a TF family which plays an important role in both the vegetative and the reproductive phases of plant life. We combine phylogenetic reconstruction, protein domain analysis, and interaction data from different species. We show that, unlike previously analyzed interaction networks, the MIKC-type protein network displays a characteristic topology, with overall high inter-subfamily connectivity, shared interactors between paralogs, and conservation of interaction patterns across species. The evaluation of the number of MIKC-type proteins at key time points throughout the evolution of land plants in the lineage leading to Arabidopsis suggested that most duplicates were retained after each round of WGD. We provide evidence that an initial network, formed by 9–11 homodimerizing proteins interacting with each other, existed in the common ancestor of all seed plants. This basic structure has been conserved after each round of WGD, adding layers of paralogs with similar interaction patterns. We thus present the first model where we can show that a network of eukaryotic TFs has evolved via rounds of WGD. Furthermore, we found that in subfamilies in which the K domain is most diverged, the interactions with other subfamilies have been largely lost. We discuss the possibility that such a high proportion of genes were retained after each WGD because of their capacity to form higher order complexes involving proteins from different subfamilies. The simultaneous duplications allowed for the conservation of the quantitative balance between the constituents and facilitated sub- and neofunctionalization through differential expression of whole units.

Introduction

Reconstructing the evolution of regulatory networks is important to understand the development of organisms and their evolutionary and physiological adaptation to environmental changes and suggests points of intervention for genetic manipulations. The abundance of genomic, proteomic, and transcriptomic data enables the reconstruction of the evolution of regulatory networks by comparative analysis of their constituent complexes. It also facilitates testing models of how complex regulatory units evolve. For example, the rewiring of genetic networks has been predicted to evolve either by series of single-gene duplications (SGDs) or whole-network duplications (Wagner 1994). Most theoretical studies concentrated on the emergence of hub-based topologies, where some proteins have many interactions and many have few, called scale-free topology (Barabási and Oltvai 2004; Yook et al. 2004). Among the most significant claims concerning network evolution were the postulates that evolutionary pressures select certain topologies (Conant and Wagner 2003; Milo et al. 2004) and that SGDs have been the most important evolutionary events in network growth (Pastor-Satorras et al. 2003; Middendorff et al. 2005), although it is widely accepted now that most eukaryotic lineages have encountered several whole-genome duplications (WGDs).

However, most of these studies have not been linked with phylogenetic analyses, and thus, the actual driving forces of network evolution remain obscure. We have previously analyzed the evolution of interaction networks in-

duced by dimerizing transcription factors, namely, the bZIP, bHLH, and NR families in animal genomes (Amoutzias et al. 2004, 2005). The main conclusions so far were that the interaction networks of these families evolved mainly by series of SGDs.

To test the possible effects of WGDs on the growth and topology of interaction networks, it is reasonable to reconstruct the evolution of such networks in plants because plants have undergone several WGDs. Of particular interest are transcription factor (TF) families such as the MADS domain proteins, which all contain the highly conserved MADS DNA-binding domain and can be found in animals, fungi, and plants. In plants, MADS domain proteins of the MIKC-type, which are widely spread, also contain I, K, and C domains in addition to the MADS domain. The K domain forms coiled-coil dimers with other K domains (Theissen et al. 2000). The C domain acts as a transcriptional activator and stabilizes protein interactions (Theissen and Saedler 2001).

MIKC-type proteins bind DNA as dimers or higher order complexes. Certain MADS protein complexes are known to regulate the expression of genes in flower meristems, whereas others act specifically in roots, flower induction, or inflorescence architecture (Theissen et al. 2000). Protein–protein interactions (PPIs) influence the function of MADS protein by affecting their DNA binding (Egea-Cortines et al. 1999) and/or transactivation potential (de Folter et al. 2005). Reports of interactions (mainly dimer formation) between MIKC-type proteins exist for many plant species. However, large-scale studies have only been performed in Arabidopsis and Petunia, using mainly the yeast-two-hybrid (Y2H) system (Immink et al. 2002; de Folter et al. 2005).

In this study, we investigate whether WGDs have had an influence on the growth and topology of the MIKC-type

Key words: genome duplication, protein network, MADS, transcription factor.

E-mail: averon@uni-muenster.de.

Mol. Biol. Evol. 24(3):670–678, 2007

doi:10.1093/molbev/msl197

Advance Access publication December 14, 2006

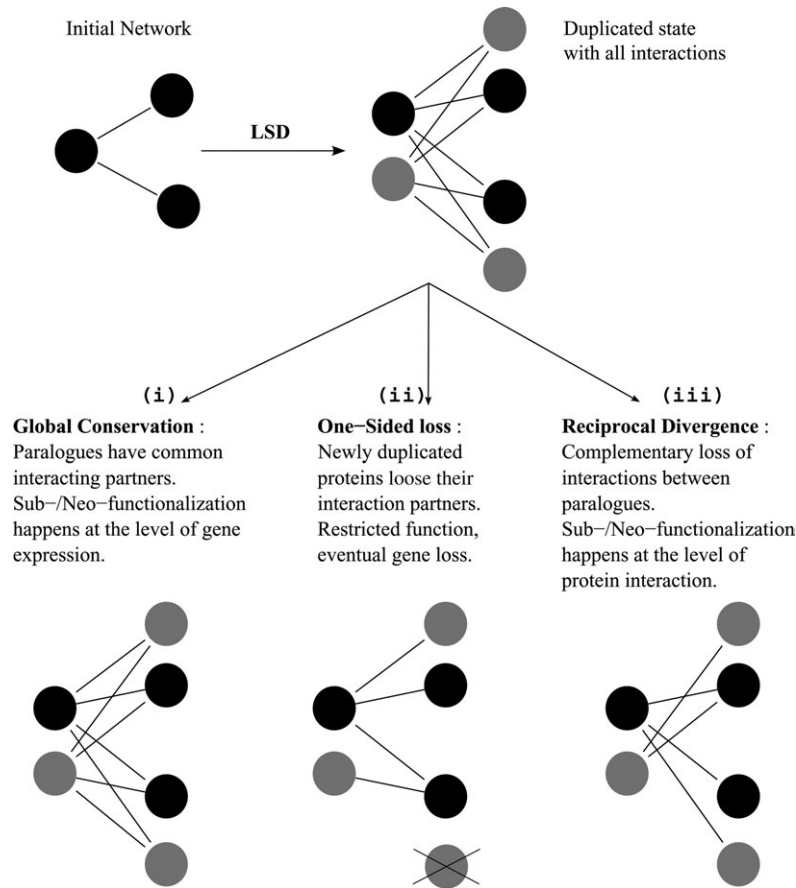


FIG. 1.—Possible scenarios of evolution following a WGD. Black circles represent ancestral proteins and gray circles represent newly duplicated proteins. A line between 2 proteins represents their ability to form heterodimers. For simplicity, only the evolution of heterodimerization is displayed.

PPI network. We propose that SGD and WGDs have different effects on the topology of a protein network. Conceptually, we first assume that WGDs, with the concomitant duplicating of all interactions, can be followed by 3 non-mutually exclusive scenarios: 1) global conservation and the appearance of redundancy between close paralogs interacting with more distant proteins; 2) one-sided loss, where a recently duplicated protein becomes poorly connected; and 3) reciprocal divergence, where the original protein also loses some interactions (see fig. 1). Global conservation and reciprocal divergence are characteristic of large-scale duplications (LSDs) because they can only occur after the simultaneous duplication of several proteins. Indeed, if only series of SGDs are used to account for the growth of a network, it becomes less probable that the topology presented in scenario 1 would be observed: several SGD events would be needed to account for this topology and all interactions of all duplicated proteins would have to be conserved over a relatively long period of time in order for the new duplicates to be able to interact both with the ancestral and with the previously duplicated proteins. Topologies of global redundancy and reciprocal divergence can therefore be considered to have arisen following WGD events.

We will first analyze the conservation of the domain architecture among subfamilies and then study the growth

of the MIKC family and the evolution of its topology. The influence of WGD events in the evolution of the MIKC-type PPI network will be discussed, as well as possible explanations for the rapid growth of this family in plants.

Materials and Methods

In this study, protein interaction refers to a physical interaction between 2 proteins, which results in the formation of a dimer. As substantial interaction data are only available for *Arabidopsis thaliana* (de Folter et al. 2005) and *Petunia hybrida* (Immink et al. 2003) and, to a lesser extent, for *Antirrhinum majus* and *Oryza sativa*, only these interaction sets were used. All interactions we used in this study were found in the literature where they have been experimentally determined. The complete lists and references are presented in supplementary tables 1–4 (Supplementary Material online).

The sequences of maize (*Zea mays*) and gnetum (*Gnetum gnemon*) were included in the phylogenetic analyses to determine the age of gene duplications (see fig. 2). Three sequences from charophycean green algae (Tanabe et al. 2005) were also added. Green algae share a common ancestor with all land plants, and thus, their sequences can serve as an outgroup to root the tree of MIKC-type genes.

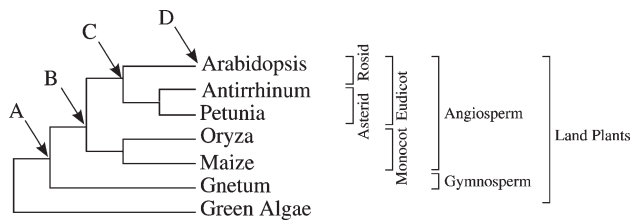


FIG. 2.—Phylogenetic relationship between the species studied. The arrows on the tree represent specific time points to which we refer in the present work. A: Angiosperm–gymnosperm split; B: eudicot–monocot split; C: rosid–asterid split; and D: present time.

The MADS, I, and K domains were defined based on an initial alignment of MIKC-type proteins. The set of retrieved sequences was then scanned for matches to the hidden Markov model of the domains. All proteins containing a MADS and/or a K domain with an *E* value less than 0.01 were selected and added to the alignment using ClustalW (Thompson et al. 1994). Only the defined domains (MADS, I, and K) were used for the maximum likelihood (ML) phylogenetic reconstruction. The visualization and annotations were done with TreeDyn (Chevenet et al. 2006) and Inkscape.

To determine the domain composition of each protein, we used the PRODOM database (Servant et al. 2002), which was constructed by automatically clustering conserved sequence fragments into families, based on multiple alignments. PRODOM domains are thus conserved sequences of amino acids across proteins but do not necessarily represent structural domains as experimentally defined. Therefore, a domain could be represented by more than one PRODOM domain ID if different portions of the domain have different degrees of conservation across proteins. The full sequence of each protein was blasted against the PRODOM database with default BlastP parameters, which returned a large number of descriptions and alignments ($-v 1000 -b 1000$). The Blast result was then processed to select between overlapping domains and to eliminate hits with an *E* value greater than 1×10^{-4} or less than 60% conserved residues.

Using the phylogeny and interaction data of today's MIKC-type protein networks, we estimated the state of the ancestral network at specific time points, just before the splits of the species and branches (see fig. 2) we considered (see fig. 4, part II). More precisely, starting from the Arabidopsis network, we estimated the network of the eudicot ancestor (i.e., just before the rosid–asterid split): for each Arabidopsis protein, if an ortholog was found in Petunia or Antirrhinum, we considered that this protein as well as its interactions was already present in their common ancestor's genome. However, if no ortholog could be found, we considered that this protein was a rosid-specific protein that was not in the ancestor's genome. We used the same deduction process to estimate the network of the common ancestor of the eudicots and monocots and the network of the ancestor of the angiosperms and gymnosperms. Whenever a protein that had interactions was found to be lineage specific, its interactions were reported to its closest conserved paralog. This is justified by the high overall conservation of interactions, which makes it more likely

that all interactions are inherited from ancestor proteins and not suddenly gained by a specific pair of proteins.

More details on the methods used can be found in the Supplementary Material online.

Results

Conservation and Divergence Patterns of Different Domains

In order to clarify the relationships between the subfamilies and of specifying changes in primary protein structure during the evolution of the MIKC-type protein family, the domain composition of the full-length proteins was determined in terms of PRODOM domains.

Globally, the MADS and K domains are under higher structural or functional constraints because they are significantly more conserved than the I and C domains (Parenicova et al. 2003). In accordance with these results, we have found that the MADS domain is highly conserved across all subfamilies (see fig. 3 and supplementary fig. 1 [Supplementary Material online]). In our analysis, the K domain is represented by several PRODOM domain IDs (see fig. 3). The first PRODOM domain (PD000423) is common to all subfamilies except the FLC subfamily and corresponds to the first 2 of the 3 predicted amphipathic α helices of the K domain (Kaufmann et al. 2005). However, the C-terminal part of the K domain (or K3) and the C domain show more variation, which is reflected by different PRODOM domains that are subfamily specific. This is in agreement with some experimental evidence that the K3 helix is less important in the formation of dimers (Yang et al. 2003) and could be under fewer structural constraints.

The SEPALLATA-like and AGL6-like subfamilies share the same PRODOM domain ID for K3 (PD352768), whereas the SQUA-like subfamily possesses a different one (see fig. 3). This indicates that SEPALLATA-like and AGL6-like subfamilies are more closely related to each other than to SQUA-like subfamily, in accordance with the phylogenetic reconstruction shown in figure 3.

Proteins of the FLC-like subfamily, which display a highly divergent K domain (see fig. 3), have not been reported to form dimers in Y2H assays and FLC homo-dimers have only been observed in gel retardation experiments, which reveals the weakness of this interaction. Similarly, the AGL12-like proteins have a (slightly less) divergent K domain, which is not predicted to form coiled-coil regions by standard tools (Alvarez-Buylla et al. 2000), and AGL12 has a very limited number of interaction partners. This peculiarity of AGL12 and FLC is also reflected at the level of their expression patterns, which are very broad in terms of number of tissues, whereas expression of most other MIKC-type proteins is more tissue specific (Becker and Theissen 2003).

The Growth of the MIKC-Type Gene Family during Seed Plant Evolution Corresponds with WGDs

Three WGDs have been detected in the angiosperm lineage leading to Arabidopsis and are estimated to have occurred in early angiosperm evolution (before the monocot–eudicot split), in early eudicot evolution (before the asterid–rosid split), and in the lineage leading to Arabidopsis

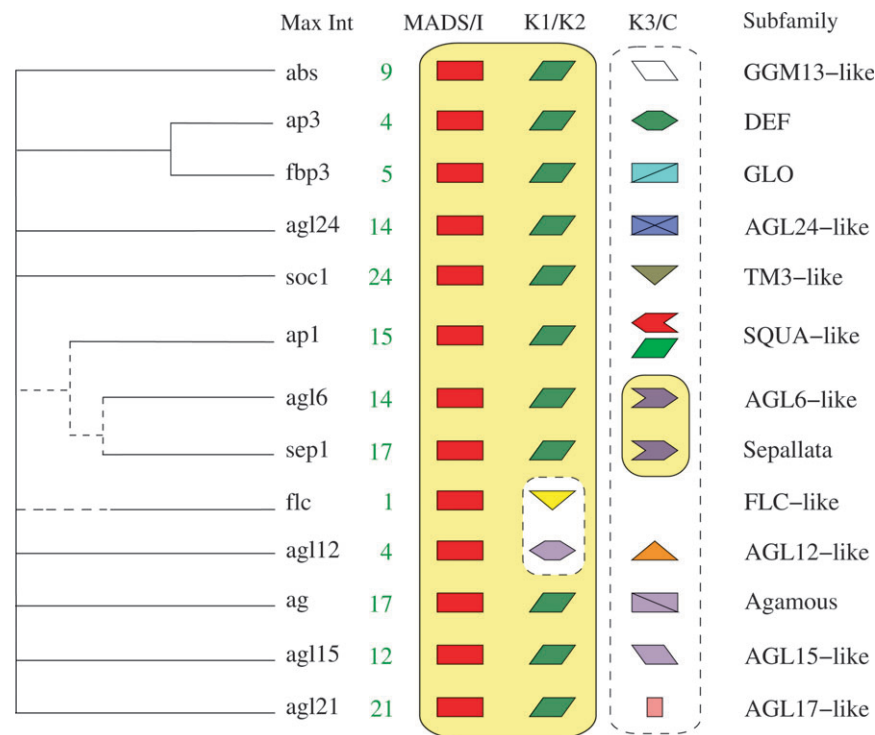


FIG. 3.—Domain conservation and divergence across the MIKC subfamilies. Different PRODOM domain IDs are represented by different symbols. Full boxes highlight similarities, and dashed boxes highlight differences. Uncertainties in the phylogeny of the MIKC-type proteins are represented by dashed lines. For each subfamily, the name and the number of interactions of the protein with the most interactions is given (Max Int). See supplementary figure 1 (Supplementary Material online) for the full tree and further discussion.

as given in (Bowers et al. 2003, de Bodt et al. 2005, Cui et al. 2006, respectively). Comparing the MIKC-type proteins of the species included in our phylogenetic analysis (see fig. 2) and inferring likely gene duplications and losses at specific time points provide insights into the size and composition of the MIKC-type protein family in ancestor organisms before each of the aforementioned splits.

As shown in the timeline of figure 4 (part I) the number of MIKC-type proteins in flowering plants has increased over time, almost doubling between each aforementioned splits, which corresponds with the timing of reported WGDs. All these duplications have occurred in most subfamilies (see the simplified phylogeny in fig. 4, part I). Therefore, although our species sampling only allows raw estimates, the timing and distribution of the duplications is consistent with the hypothesis that WGDs had a major role in the formation of the MIKC-type protein family. We next assess whether this series of LSDs also had an influence on the PPI network of the MIKC-type proteins.

An Ancestral Network of 9–11 Proteins Is Still the Core of the MIKC-Type Protein Network

We estimated the number of MIKC-type proteins of the seed plant ancestor (just before the gymnosperm–angiosperm split) to be between 9 and 11: in our phylogenetic reconstruction (see supplementary fig. 1, Supplementary Material online), only 8 subfamilies possess a gymnosperm ortholog, but GGM2 has been reported to belong to the DEF-GLO subfamily, bringing the number to 9 (Winter

et al. 2002). Additionally, it is possible that the respective ancestral protein of the SQUA and SEPALLATA subfamilies already existed in the genome of the early seed plant ancestors (Zahn et al. 2005). According to our modeling of the ancestral interaction patterns, the overall topology of the network has remained nearly constant during evolution (see fig. 4, part II, B–D). Therefore, according to this model, the 3 rounds of WGDs have added layers of interacting paralogs without modifying the original topology of the MIKC-type PPI networks, which stretches back to before the angiosperm–gymnosperm split.

The MIKC Interaction Network Has a Specific Topology

We find 3 specific properties of the topology of the MIKC-type protein interaction network, which distinguish it from other eukaryotic networks (Amoutzias et al. 2004). Experiments involving loss-of-function mutants have shown that paralogous MIKC-type proteins had partially redundant roles as triple or even quadruple mutants were required to observe particular phenotypic effects (Pelaz et al. 2001; Pinyopich et al. 2003). Here we provide an explanation for this observed functional redundancy by revealing how frequently paralogous proteins share similar interaction partners. In the SEPALLATA subfamily, for example, we discovered that within each species, all the paralogs have a very similar interaction pattern (see fig. 5). This is particularly striking in *Petunia*, where the 6 SEPALLATA-like paralogs have similar interaction patterns, 4 of which are able to form homodimers, of which 3 are also able to form

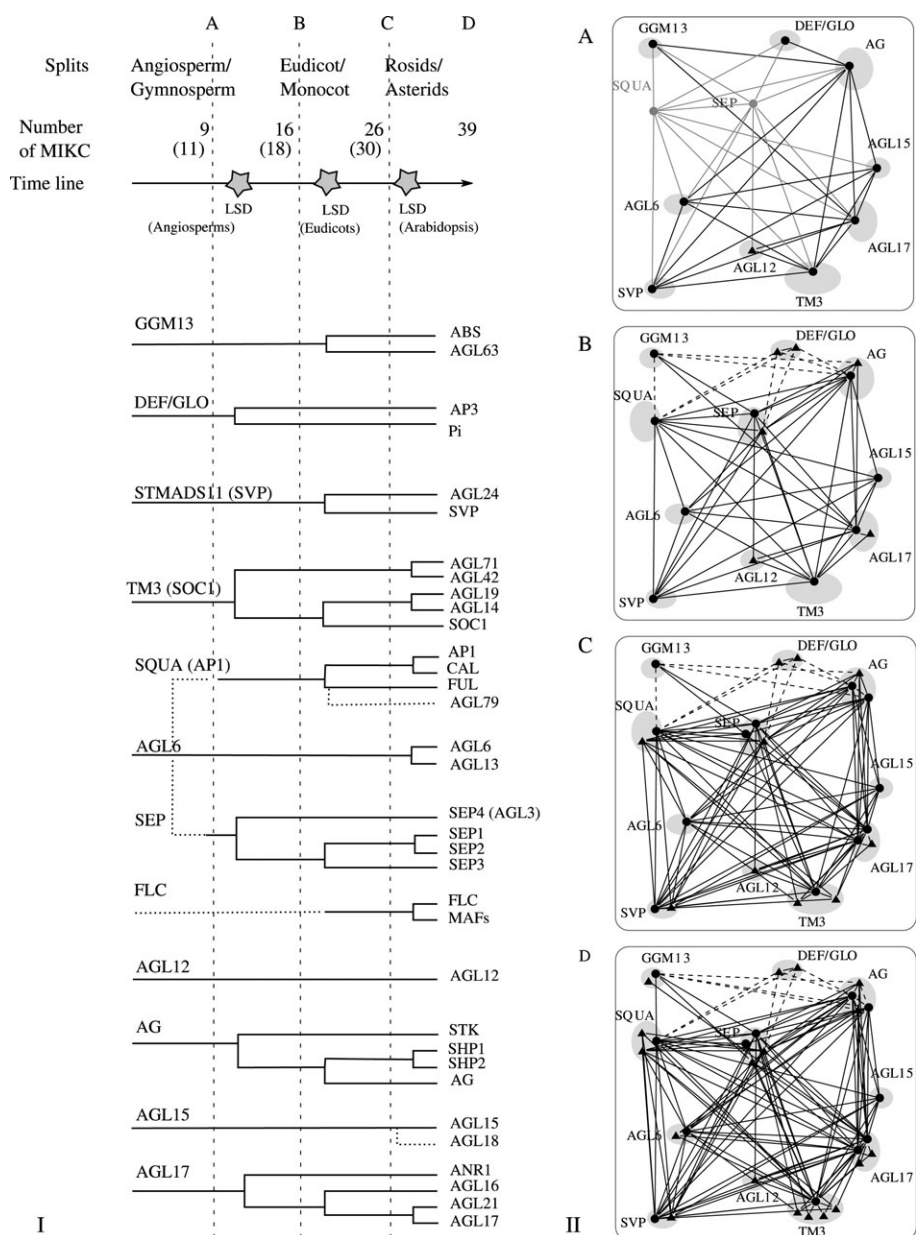


FIG. 4.—Evolutionary model for the MIKC-type protein network. Only Arabidopsis proteins are shown. Part I displays a simplified phylogeny of the MIKC-type proteins, where dotted lines represent uncertainties. A timeline is presented at the top of the figure on which stars represent the 3 most recent LSD events (de Bodt et al. 2005), and vertical dotted lines indicate splits. The estimated minimum number of MIKC-type proteins before each of the aforementioned splits along with the maximum number in parentheses (uncertainties due to occasional low bootstrap values in the tree presented in supplementary fig. 1, Supplementary Material online) are also displayed. Part II shows estimations of the MIKC-type PPI network at different times. (A) Before the angiosperm–gymnosperm split, (B) before the eudicot–monocot split, (C) before the rosid–asterid split, and (D) today. Black circles represent homo- and heterodimerizing proteins, triangles represent heterodimerizing-only proteins, full lines represent dimerization, and dotted lines represent interactions observed only in higher order complexes.

higher order complexes with the same partners. These proteins therefore display a high degree of global redundancy, as defined in figure 1 (i). In Arabidopsis, 3 of the 4 SEPALLATA paralogs have been observed to form homodimers and share the same interaction partners, with some level of reciprocal divergence (see fig. 1, iii). In the lesser studied species *Antirrhinum* and *Oryza*, the data we collected seemed to confirm this tendency of PPI redundancy among paralogs (see fig. 5). Interestingly, some of the SEPALLATA paralogs originated from a duplication preceding

the eudicot–monocot split, which proves that even distant duplicates can have similar interaction patterns. What is striking about the conservation of interactions between paralogous proteins in the MIKC-type network is that this conservation is observed across all subfamilies (see supplementary figs. 2–5, Supplementary Material online).

In all subfamilies for which interaction data are available, the interaction patterns of orthologs are strikingly similar. For example, in Arabidopsis, SEP3 (a SEPALLATA-like protein) interacts with SHP1 (of the AGAMOUS

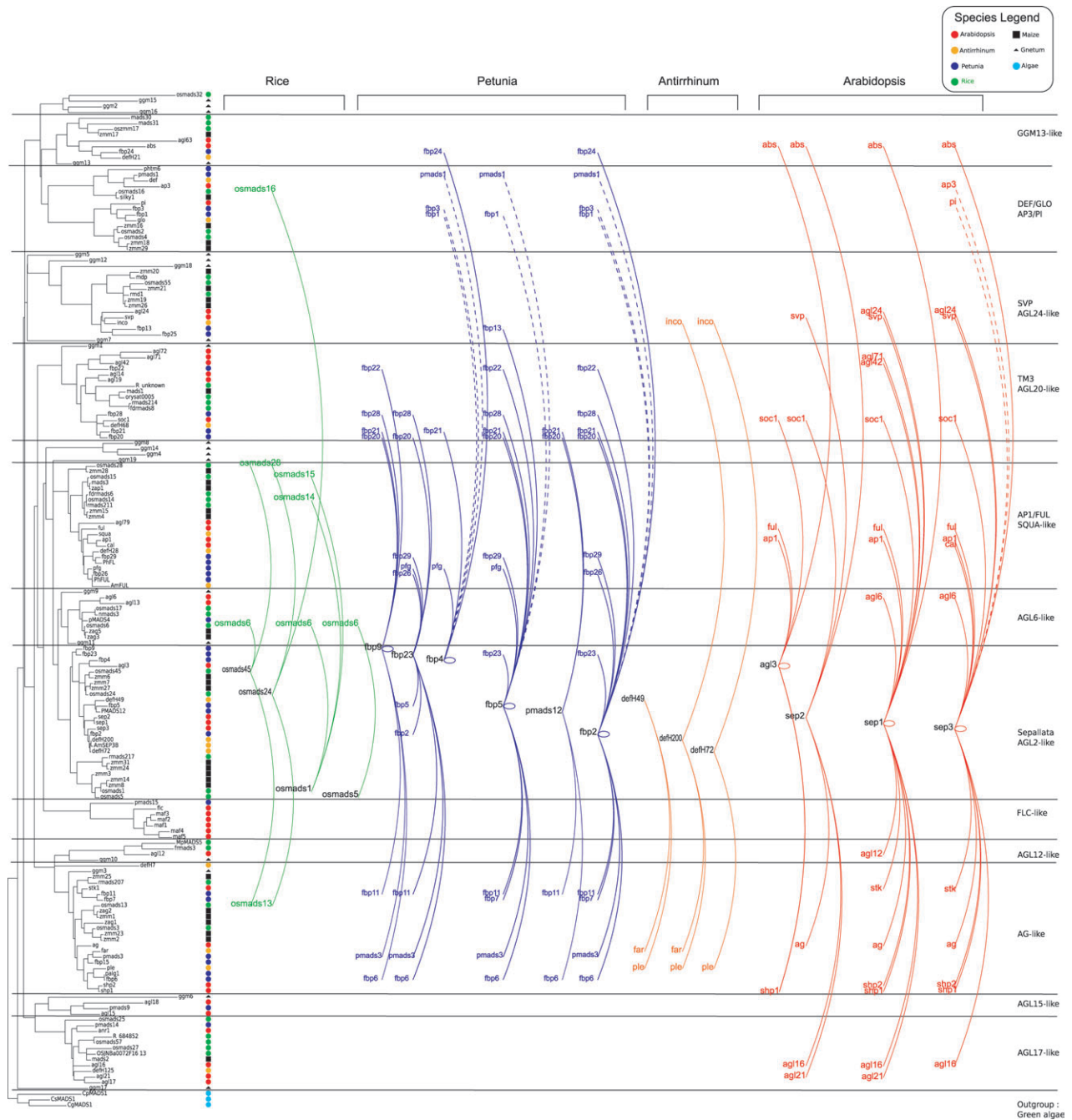


FIG. 5.—An example of the topology of the MIKC network: the MADS-box SEPALLATA-like PPI display high intra- and interspecies conservation as well as high inter-subfamily connectivity. The phylogenetic tree is the same ML tree as presented in supplementary figure 1 (Supplementary Material online). The topology presented here is a global property of the MIKC network, as can be seen in supplementary figures 2–5 (Supplementary Material online). For each protein of the SEPALLATA subfamily for which interactions are known, the corresponding interactions are displayed in the form of a curve linking the protein to its diverse interaction partners. Whereas dimerizations are represented by continuous lines, interactions only observed in higher order complexes are represented by discontinuous lines.

subfamily), whereas in *Petunia*, FBP2 and FBP6, which are direct orthologues of SEP3 and SHP1, respectively, also interact (see fig. 5). This conservation of interaction patterns is found throughout the available data. There are still numerous putative protein interactions that have not been investigated, such as the interactions of the AGL17-like proteins, for which only *Arabidopsis* data are available (see supplementary fig. 2, Supplementary Material online).

Similarly, available data show that higher order complex formation is conserved across species as well as homodimerization. This striking conservation of interaction patterns across species is likely to be even stronger as more interaction data are collected.

Another specific property of the MIKC-type protein interaction network is that the majority of the heterodimers that they form consists of proteins from different subfamilies

(see figs. 4 [part II, *D*] and 5). When we consider 2 subfamilies as interacting if there is at least one protein of a subfamily that interacts with a member of the other subfamily, we find that of a total of 12 subfamilies, 6 formed a clique. In graph theory, a clique in an undirected graph G is a set of vertices V such that for every 2 vertices in V , there exists an edge connecting the two. Additionally, 2 more subfamilies could be added to this set of highly connected subfamilies, each of them missing only one interaction: these 8 subfamilies thus form 24 of the 26 possible interactions. However, not all MIKC subfamilies display the same connectivity: 4 subfamilies (sister clades GGM13 and DEF-GLO and highly diverged K domain AGL12 and FLC) are noticeably less connected (see fig. 4, part II, *D*).

Discussion

Our results support the theory that the expansion of the MIKC-type gene family was mainly due to WGDs. Moreover, we found that as a general rule in the MIKC-type PPI network, the interactions of paralogous proteins were highly similar, orthologs had conserved interaction patterns, and the subfamilies were highly connected by PPI. These additional results reveal the unique topology of the MIKC-type PPI network.

Indeed, by comparing the MADS MIKC-type protein network with other TF networks of similar size—bHLH, NR, and bZIP (Amoutzias et al. 2004, 2005, 2007)—we identified intriguing common principles between these eukaryotic networks and specificities of the MADS network. In all 4 networks, we observed that heterodimerizing proteins emerged via duplication and divergence of homodimerizing ancestors. In the MADS just as in the bHLH and the NR protein interaction networks, the gain of additional domains—here the I, K, and C domains which were gained before the split of land plants and charophycean green algae (Tanabe et al. 2005)—has influenced the topology of the network, causing compartmentalization of the interactions, which is reflected by the fact that most interactions happen between MIKC-type proteins and not with other types of MADS proteins (de Folter et al. 2005). The level of inter-subfamily connectivity, on the other hand, is strikingly more important in the MIKC network than in the bHLH network, where interactions between distantly related family members are uncommon. In addition, the level of conservation of interaction partners, both between paralogs and across species, is particularly high in the MIKC-type protein network.

We therefore suggest that the relatively fast succession of WGDs, which have occurred in angiosperm evolution, have influenced not only the growth of the MIKC PPI network but also its topology. The fact that so many interactions are repeated between paralogs and orthologs makes it very likely that these interactions are inherited and conserved from the ancestral proteins, after duplication and/or speciation. Based on this hypothesis and the deductive process presented in Materials and Methods, we suggest an evolutionary scenario for the MIKC-type PPI network, which is presented in figure 4 (part II). This network growth model of LSD followed by restricted loss of gene and PPI

could explain the formation of other protein networks in protein classes known to be preferentially retained after WGD events (Maere et al. 2005).

However, given the degree of redundancy of interaction patterns among paralogs, how can the retention of so many MIKC-type genes be explained, when most of the other classes of duplicated genes are lost after WGD events? One of the distinguishing features of the MIKC-type proteins is their capacity to form higher order complexes, the composition of which is conserved across species. These complexes, which have been isolated in all eudicot species considered in this study, are known to bind DNA *in vivo* and are essential for the formation of flowers in angiosperms (Honma and Goto 2001; Pelaz et al. 2001). Considering the importance of these complexes for the regulation of essential plant organs, the simultaneous duplication of all constituents could have presented a considerable advantage, allowing for a higher number of such complexes and their differential expression (Kaufmann et al. 2005).

Considering the conservation of the network topology across species and between paralogs, it is likely that differential phenotypes are induced by differential expression of paralogs during development, selective DNA binding, and differential transcription activation capacity (Ditta et al. 2004). Gene duplications and the differential expression of the resulting paralogs enable the fine-tuning of gene regulation in different tissues or under different conditions. This is reflected by the fact that in ferns the expression of MIKC genes is more ubiquitous than in seed plants (Münster et al. 1997).

Therefore, the fact that so many MIKC-type proteins have been retained along with conserved interaction patterns can be at least partially explained by their differential expression.

Conclusion

The integrated analysis of sequence, phylogenetic, and PPI data has revealed the specific structure of the MIKC network, characterized by a high number of shared interactors between paralogs, high interaction pattern conservation between orthologs, and high inter-subfamily connectivity. We have compared this topology with that of other networks of similar size and delineated common principles and specific traits of the MADS network. Together with the estimated duplication time of the MIKC-type genes and their homogenous distribution among most subfamilies, these results clearly indicate that series of WGDs shaped the structure of the MIKC network.

The question of why MIKC genes were specifically retained after WGD events is still open. However, one possible advantage of the duplicates is the increased combinatorial potential for multimer formation, allowing for differential expression of paralogous proteins and complex-specific DNA-binding properties. Finally, the simultaneous expansion of pollinators and angiosperms approximately 65 MYA supports the hypothesis that pollinators played a key role in the evolution of flowering plants and vice versa. As the MADS-box proteins have a direct role in the formation of reproductive organs, it is probable that the expansion and

topology of the MADS PPI network has facilitated this co-adaptation.

Supplementary Material

Supplementary tables 1–4 and figures 1–5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We are grateful to Francois Chevenet who updated the TreeDyn software and to January Weiner for providing us with perl scripts. We would also like to thank Klaus Harter, Stefanie de Bodt, Jeorg Kudla, and Thorsten Reusch for insightful comments and suggestions on the manuscript and Jan-Michael Rye for proofreading it. A.V. and E.B.-B. acknowledge support from Deutsche Forschungsgemeinschaft through project grant number BO-2544/2-1. A.V. was also supported by a Marie Curie funding.

Literature Cited

- Alvarez-Buylla ER, Pelaz S, Liljegen SJ, Gold SE, Burgeff C, Ditta GS, de Poupiana LR, Martinez-Castilla L, Yanofsky MF. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA*. 97:5328–5333.
- Amoutzias G, Robertson D, Oliver S, Bornberg-Bauer E. 2004. Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. *EMBO Rep*. 5:274–279.
- Amoutzias G, Veron AS, Weiner III J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, Robertson DL. 2007. One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Mol Biol Evol*. 10.1093/molbev/msl211.
- Amoutzias G, Weiner J, Bornberg-Bauer E. 2005. Phylogenetic profiling of protein interaction networks in eukaryotic transcription factors reveals focal proteins being ancestral to hubs. *Gene*. 347:247–253.
- Barabási A-L, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 5:101–113.
- Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*. 29:464–489.
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*. 422:433–438.
- Chevenet F, Brun C, Bauls A-L, Jacq B, Christen R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinform*. 7:439.
- Conant G, Wagner A. 2003. Convergent evolution of gene circuits. *Nat Genet*. 34:264–266.
- Cui L, Wall PK, Leebens-Mack JH, Doyle JJ, Soltis DE, Soltis PS, Carlson J, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res*. 16:738–749.
- de Bodt S, Maere S, de Peer YV. 2005. Genome duplication and the origin of angiosperms. *Trends Ecol Evol*. 20:591–597.
- de Folter S, Immink R, Kieffer M, Parenicova L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, et al. (12 co-authors). 2005. Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *Plant Cell*. 17:1424–33.
- Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004. The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr Biol*. 14:1935–1940.
- Egea-Cortines M, Saedler H, Sommer H. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *EMBO J*. 18:5370–5379.
- Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*. 409:525–529.
- Immink RGH, Ferrario S, Busscher-Lange J, Kooiker M, Busscher M, Angenent GC. 2003. Analysis of the petunia MADS-box transcription factor family. *Mol Genet Genomics*. 268:598–606.
- Immink RGH, Gadella TWJ, Ferrario S, Busscher M, Angenent GC. 2002. Analysis of MADS box protein-protein interactions in living plant cells. *Proc Natl Acad Sci USA*. 99:2416–2421.
- Kaufmann K, Melzer R, Theissen G. 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene*. 347(2):183–198.
- Maere S, Bodt SD, Raes J, Casneuf T, Montagu MV, Kuiper M, de Peer YV. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci USA*. 102:5454–5459.
- Middendorff M, Ziv E, Wiggins C. 2005. Inferring network mechanisms: the *Drosophila melanogaster* protein interaction network. *Proc Natl Acad Sci USA*. 102:3192–3197.
- Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U. 2004. Superfamilies of evolved and designed networks. *Science*. 303:1538–1542.
- Münster T, Pahnke J, Rosa AD, Kim JT, Martin W, Saedler H, Theissen G. 1997. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc Natl Acad Sci USA*. 94:2415–2420.
- Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, et al. (11 co-authors). 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*. 15:1538–1551.
- Pastor-Satorras R, Smith E, Sol RV. 2003. Evolving protein interaction networks through gene duplication. *J Theor Biol*. 222:199–210.
- Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF. 2001. APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J*. 26:385–394.
- Pinyopich A, Ditta GS, Savidge B, Liljegen SJ, Baumann E, Wisman E, Yanofsky MF. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*. 424:85–88.
- Servant F, Bru C, Carrre S, Courcelle E, Gouzy J, Peyruc D, Kahn D. 2002. ProDom: automated clustering of homologous domains. *Brief Bioinform*. 3:246–251.
- Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitani M, Henschel K, Munster T, Theissen G, Nozaki H, Ito M. 2005. Characterization of MADS-box genes in charophyte green algae and its implication for the evolution of MADS-box genes. *Proc Natl Acad Sci USA*. 102:2436–2441.
- Theissen G, Becker A, Rosa AD, Kanno A, Kim JT, Muenster T, Winter KU, Saedler H. 2000. A short history of MADS-box genes in plants. *Plant Mol Biol*. 42:115–149.
- Theissen G, Saedler H. 2001. Plant biology. Floral quartets. *Nature*. 409:469–471.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
- Wagner A. 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci USA*. 91:4387–4391.

- Winter K-U, Weiser C, Kaufmann K, Böhne A, Kirchner C, Kanno A, Saedler H, Theissen G. 2002. Evolution of class B floral homeotic proteins: obligate heterodimerization originated from homodimerization. *Mol Biol Evol.* 19:587–596.
- Yang Y, Fanning L, Jack T. 2003. The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA. *Plant J.* 33:47–59.
- Yook S-H, Oltvai ZN, Barabási A-L. 2004. Functional and topological characterization of protein interaction networks. *Proteomics.* 4:928–942.
- Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, Depamphilis CW, Ma H. 2005. The evolution of the SEPALLATA subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics.* 169:2209–2223.

William Martin, Associate Editor

Accepted December 5, 2006