

## Evolution of the *RALF* Gene Family in Plants: Gene Duplication and Selection Patterns

Jun Cao<sup>1</sup>, and Feng Shi<sup>2</sup>

<sup>1</sup>Institute of Life Science, Jiangsu University, Xuefu Road 301, Zhenjiang (212013), Jiangsu, PR China. <sup>2</sup>Shandong Lvdu Bio-technique Industry, 169# Huanghe 2 Road, Binzhou (256600), Shandong, PR China.  
Corresponding author email: [cjinfor@163.com](mailto:cjinfor@163.com)

---

**Abstract:** Rapid alkalization factors (RALFs) are plant small peptides that could induce a rapid pH increase in the medium of plant cell suspension culture and play a critical role in plant development. The evolutionary process of the *RALF* gene family remains unclear. To obtain details of the phylogeny of these genes, this study characterized *RALF* genes in *Arabidopsis*, rice, poplar and maize. Phylogenetic trees, evolutionary patterns and molecular evolutionary rates were used to elucidate the evolutionary process of this gene family. In addition, the different signatures of selection, expression patterns, and subcellular localization of *RALFs* were also analyzed. We found that the *RALF* gene family had a rapid birth process after the separation of the eudicot and monocot species about 145 million years ago, that tandem duplication played a dominant role in the expansion of *Arabidopsis* and rice *RALF* gene family, and that *RALFs* were under purifying selection according to estimations of the substitution rates of these genes. We also identified a diverse expression pattern of *RALF* genes and predominant extracellular localization feature of RALF proteins. Our findings shed light on several key differences in *RALF* gene family evolution among the plant species, which may provide a scaffold for future functional analysis of this family.

**Keywords:** *RALF*, tandem duplication, evolution, selection

---

*Evolutionary Bioinformatics* 2012:8 271–292

doi: [10.4137/EBO.S9652](https://doi.org/10.4137/EBO.S9652)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



## Introduction

Peptide signaling is important for cell-to-cell communication and participates in a variety of developmental processes and environmental responses. A number of genes encoding small-secreted peptides have been identified in plants and a certain proportion of them are hormones.<sup>1</sup> These peptides play critical role in all aspects of the plant life cycle and have diverse functions. Such as, CLV3 (CLAVATA3) peptide regulates meristem size.<sup>2</sup> Peptide systemin induces the systemic defense response.<sup>3</sup> ENOD40 encodes two small peptides, both of which can affect the normal nodule development.<sup>4</sup> Defensins are involved in the innate immune system of plants.<sup>5</sup> PSK (phytosulfokine) peptide has been demonstrated to promote cellular proliferation and transdifferentiation.<sup>6,7</sup> SCR peptide is the pollen self-incompatibility recognition factor in the *Brassicaceae* species.<sup>8,9</sup> PLS (POLARIS) peptide is involved in vascularization, longitudinal cell expansion and increased radial expansion.<sup>10</sup> ROT4/DVL1 (ROTUNDIFOLIA4/DEVIL1) peptide regulates polar cell proliferation on the longitudinal axis of organs.<sup>11,12</sup> IDA (INFLORESCENCE DEFICIENT IN ABSCISSION) is a family of secreted peptides identified to be involved in petal abscission.<sup>13</sup> LURE peptides produced by synergid cells attract pollen tubes to the embryo sac.<sup>14</sup> RGF (root meristem growth factor) is a 13 amino acid secreted peptide involved in the maintenance of root stem cell niche.<sup>15</sup> Dodeca-CLE peptides suppress the plant stem cell differentiation.<sup>16</sup> In addition, RALF is a recently discovered family of plant peptide that plays a role in plant cell growth as will be described below.<sup>17</sup>

*RALF* is a small peptide and first discovered in tobacco leaf extracts due to its ability to cause a rapid alkalinization in the medium of tobacco cell suspension cultures.<sup>18</sup> Subsequently, this gene is also identified in a wide variety of plant species, including gymnosperms, monocots and dicots.<sup>19–27</sup> The ubiquity of *RALF* suggests its importance in plant growth and development. Like other plant polypeptide hormones, such as phytosulfokine<sup>6</sup> and systemin,<sup>28</sup> most *RALF* genes encode pro-peptides that need proteolytic processing. These RALF precursors have a conserved dibasic site upstream of the active peptide that is required for pro-peptide processing and activity.<sup>29,30</sup> These results are consistent

with localization of the *Nicotiana benthamiana* RALF-GFP fusion protein, which localizes first to the endoplasmic reticulum (ER) and later to the cell wall.<sup>31</sup> Another characteristic of the peptide is the four conserved cysteines in the active peptide region that are likely to be involved in disulfide bridges and are required for activity.<sup>18</sup>

RALFs are a new type of plant peptide hormones that participate in diverse biological processes. Such as, activation of protein kinases, inhibition of root growth and development,<sup>18,21,27</sup> regulation of fruit maturation,<sup>20</sup> nodule formation,<sup>22</sup> tissue expansion<sup>10</sup> and pollen development,<sup>24,26</sup> and so on. Interestingly, the number of *RALF* genes varies greatly from species to species. For instance, over 30 *RALF* genes have been identified in *Arabidopsis*,<sup>32</sup> while only two *RALF* genes have been confirmed in *Selaginella moellendorffii* (see below). The critical role of RALFs and the diversity of *RALF* gene number from species to species prompt us to investigate how *RALF* genes have evolved in plant kingdom and how and why different species have acquired such different numbers of *RALF* gene. Here, we presented evidence that the evolution of the plant *RALF* gene family had a rapid birth process, and that tandem duplication rather than segmental duplication played a dominant role in the expansion of the *RALF* gene family. Our study also revealed a diverse expression pattern of *RALF* genes and the predominant extracellular localization feature of RALF proteins in *Arabidopsis*.

## Materials and Methods

### Sequence identification and conserved motif analysis of *RALF* genes

To identify potential members of the *RALF* gene family in *Arabidopsis*, poplar, rice and maize, we performed multiple database searches. Published *Arabidopsis* *RALF* gene sequences<sup>32</sup> were retrieved and used as queries in BLAST searches against NCBI (<http://www.ncbi.nlm.nih.gov>) and Phytozome (<http://www.phytozome.net>). The program MEME<sup>33</sup> (<http://meme.sdsc.edu>) was used to identify motifs in the candidate RALF protein sequences. MEME was run locally with the following parameters: number of repetitions = any, maximum number of motifs = 6, and with optimum motif widths constrained to between 6 and 200 residues.



## Alignment and phylogenetic analysis

To generate the alignment of the 91 RALF proteins from the *Arabidopsis*, rice, poplar and maize, COBALT<sup>34</sup> program was used. Phylogenetic analyses of the RALF proteins based on amino acid sequences were carried out using Neighbor-Joining (NJ) methods in MEGA 5.<sup>35</sup> NJ analyses were done using *p*-distance methods, pairwise deletion of gaps, and the default assumptions that the substitution patterns among lineages and substitution rates among sites were homogeneous. Support for each node was tested with 1,000 bootstrap replicates.

## Estimation of the maximum number of gained and lost RALFs

To determine the degrees of gene family expansion in the analyzed plant lineages, we divided the phylogeny into ancestral clades (those containing at least one representative of monocots and eudicots), recent clades (monocot specific or eudicot specific) and species-specific clades. Nodes basal to the split among lineages denoted the most recent common ancestor (MRCA) and were labeled as N0 to N4. Gene duplication and loss events were inferred by reconciling the gene tree for each cluster/subcluster with the species tree using Notung v2.5.<sup>36</sup>

## Divergence levels analysis

To analyze positive or negative selection of the RALF sequences, substitution rate ratios of nonsynonymous ( $K_a$ ) versus synonymous ( $K_s$ ) mutations were calculated. We first identified the closest orthologs for each gene in the genome of the close relative *A. lyrata* (Fig. S6) and included only those *A. thaliana* genes that had a single ortholog in *A. lyrata*. Moreover, gene pairs were considered orthologs when they clearly formed a single subclade. Pairwise alignment of nucleotide sequences of the RALF orthologs was performed using MEGA 5.<sup>35</sup> Alignments were performed using Clustal W (Codons).  $K_a$  and  $K_s$  values of the orthologous genes were estimated using K-Estimator 6.0.<sup>37</sup> To calculate the  $K_a/K_s$  ratios in different Groups, the Selecton server<sup>38,39</sup> was also used. It implements several evolutionary models that describe, in probabilistic terms, how characters evolve. The models are expressive enough to describe the biological reality. In this study, five models (M8, M8a, M7, M5 and MEC) were used. Each of the models uses different

biological assumptions so that different hypotheses can be tested.

## Inference of duplication time

Pairwise alignment of nucleotide sequences of the RALF paralogs was performed using MEGA 5.<sup>35</sup> Alignments were performed using Clustal W (Codons). The  $K_a$  and  $K_s$  values of the paralogous genes were estimated using K-Estimator 6.0.<sup>37</sup> To better explain the patterns of macroevolution, estimates of the evolutionary rates were considered extremely useful. Assuming a molecular clock, the synonymous substitution rates ( $K_s$ ) of the paralogous genes will be expected to be similar over time. Thus,  $K_s$  can be used as the proxy for time to estimate the dates of the segmental duplication events. The  $K_s$  value was calculated for each of the gene pairs and then used to calculate the approximate date of the duplication event ( $T = K_s/2\lambda$ ), assuming clock-like rates ( $\lambda$ ) of synonymous substitution of  $1.5 \times 10^{-8}$  substitutions/synonymous site/year for *Arabidopsis*,<sup>40</sup>  $6.5 \times 10^{-9}$  for rice and maize<sup>41</sup> and  $9.1 \times 10^{-9}$  for poplar<sup>42</sup>.

## Codon bias analysis

Codon bias can reflect the degree of selective constraint in a gene. To measure the extent of codon bias, effective number of codons (ENC) and codon bias index (CBI) were estimated using DnaSP v.5.10.01.<sup>43</sup> The ENC values range from 20 to 61, meaning from the maximum codon bias (only one codon is used for each amino acid) to no codon bias (all synonymous codons for each amino acid are equally used).<sup>44</sup> The CBI values range from 0 to 1, meaning from uniform use of synonymous codons to maximum codon bias.<sup>45</sup> We also estimated some parameters related to codon bias, such as GC1,2 (the GC content at the first and second codon positions), GC3 (the GC content at the third codon positions) using DnaSP v.5.10.01.<sup>43</sup>

## Correlation analysis of expression data and protein subcellular localization

Expression profiling can provide useful clues to gene function. To examine the expression patterns of the RALF genes, a comprehensive expression analysis was performed using the publicly available microarray data from Genevestigator.<sup>46,47</sup> For genes with more than one set of probes, the median of expression values was used. Finally, the expression data were



gene-wise normalized and hierarchically clustered based on Pearson coefficients with average linkage in the Genesis (version 1.7.6) program.<sup>48</sup> Protein subcellular localization was predicted using WoFL PSORT software (<http://wolfsort.org>).<sup>49</sup>

## Results and Discussion

### Identification, motif organization and phylogenetic analyses of the *RALF* genes

We identified 33, 23, 16, and 19 putative *RALF* genes from *Arabidopsis*, poplar, rice and maize, respectively. *Arabidopsis* has about doubled the collection of *RALF* genes than rice, whereas poplar and maize have fewer (30.3% and 42.4%, respectively) genes than *Arabidopsis*. By searching the PlantGDB (<http://www.plantgdb.org>),<sup>50</sup> we found that the predicted genomes of poplar, rice and maize contain 45,778, 30,192 and 32,540 genes, respectively, which are 67.2%, 10.3% and 18.8% larger than that of *Arabidopsis* (27,379), respectively. This suggested that the numbers of the *RALF* genes are not proportional to the sizes of the predicted genomes. All the *RALFs* in the four species possess only one *RALF* domain through the CDD<sup>51,52</sup> and Pfam (<http://pfam.sanger.ac.uk>) analysis. While these tools are suitable for defining the presence or absence of recognizable domains, they are unable to recognize smaller individual motifs and more divergent patterns. Thus, we further used the MEME program<sup>33</sup> to study the diversification of *RALF* genes in *Arabidopsis*, poplar, rice and maize. Six distinct motifs were identified in these genes (Table 1 and Fig. S1). Details of the six motifs were presented in Fig. S2.

Phylogenetic analyses can allow us to identify evolutionarily conservative and divergent of gene family. To achieve this goal, phylogenetic analyses of the 91 *RALF* members were performed. Based on phylogenetic relationships, we divided the *RALF* members into 10 groups (Fig. S1). Most of these genes encode proteins with the same or similar motif organization, while others are scattered in the families formed by proteins with other motifs, suggesting their complex evolutionary history. For convenience, we categorized the ortholog clades into 3 classes: (i) superstable: clades with orthologs containing at least one representative of monocots and eudicots, (ii) stable: clades including orthologs with monocot

**Table 1.** Number and motif structure of *RALF* proteins from *Arabidopsis* (At), poplar (Pt), rice (Os) and maize (Zm).

Group	At	Pt	Os	Zm	Structure
I	15	1	0	0	Motif 4-1 / Motif 5-1
II	7	0	0	0	Motif 5-6-1
III	0	0	3	2	Motif 5-1
IV	0	4	0	1	Motif 4-1 / Motif 5-1
V	1	5	0	0	Motif 5-3-1 / Motif 3-1 / Motif 5-3-2-1
VI	2	0	0	0	Motif 5-3-1
VII	0	0	10	13	Motif 5-2-1 / Motif 5-3-2-1 / Motif 5-6-2-1 / Motif 3-2
VIII	1	3	0	0	Motif 5-3-2-1 / Motif 5-2-1 / Motif 2-1
IX	1	2	3	3	Motif 5-2-1 / Motif 5-3-2-1 / Motif 2-1
X	6	8	0	0	Motif 5-3-2-1 / Motif 3-2-1 / Motif 2-1 / Motif 3

**Note:** Detailed illustration of the six motif structures are shown in Figure S2.

specific or eudicot specific, and (iii) unstable: lineage-specific clades. From Figure S1, it was clear that the superstable clade (Group IX) contained similar numbers of genes from each species, suggesting that major expansion/contraction in gene number had not occurred since the divergence between eudicots (*Arabidopsis* and poplar) and monocots (rice and maize). This result was also consistent with the number of *RALF* genes in *Selaginella moellendorffii*, in which only two *RALF* genes were found (Table S1). Figure S1 also showed that some genes formed lineage-specific clusters. The largest of such cluster had seven *Arabidopsis* genes. Moreover, of 16 *RALF* genes in Group I, 15 genes came from *Arabidopsis*. All of these suggested that many subsets of the *RALF* gene family had experienced extensive gene duplications.

### Contrasting changes in the numbers of *RALF* genes

To better understand how *RALF* genes have evolved in these species, we estimated the number of *RALF* genes in the MRCA of eudicots and monocots. Reconciliation of the gene trees with the species phylogeny suggested that there were about two ancestral *RALF* genes in the MRCA of eudicots and monocots (N1). Furthermore, we identified 5 orthologous genes in the eudicots MRCA (N2) and



11 in the MRCA of monocots (N3) (Fig. 1). We also found that the number of *RALFs* remained relatively stable through evolutionary history from the land plants (N0, *Physcomitrella patens*) to the vascular plants (N0, *Selaginella moellendorffii*) and the angiosperms (N1). Only after the separation of the eudicot and monocot species about 145 million years ago<sup>53</sup> did *RALFs* once more expand significantly. When compared the number of ancestral genes with those in the extant species, it appeared that the *RALF* family had expanded in all the analyzed species. For example, the number of *RALFs* increased approximately 6.6-fold since the divergence of the various eudicot species from their respective MRCA in *Arabidopsis*. However, the expansion was uneven between these plant species. For example, there are 33, 23, 16 and 19 genes in *Arabidopsis*, poplar, rice and maize, respectively, while the estimated number of genes in the MACA of eudicots and monocots are two. Therefore, *Arabidopsis*, poplar, rice and maize have gained 31, 22, 14 and 17 genes, respectively, since their splits. Only one lost gene is found in poplar. Clearly, the numbers of genes gained in the

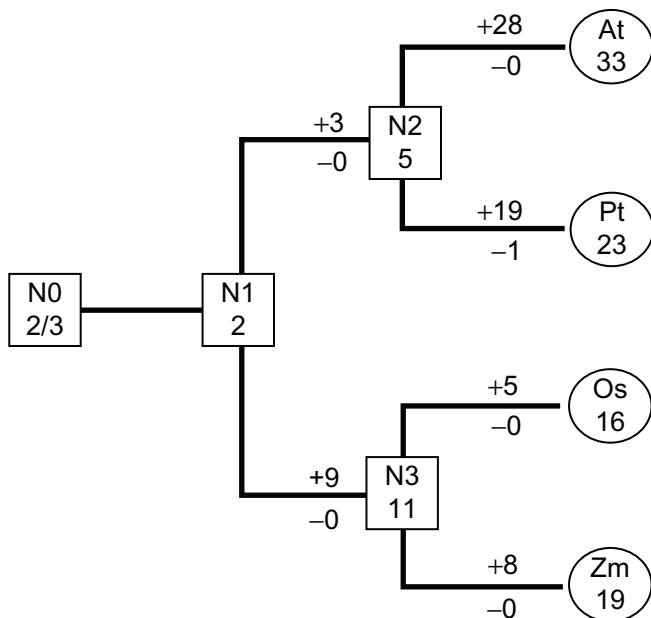
*Arabidopsis* lineage are much greater than that in other three lineages.

### Evolutionary patterns of *RALF* gene family

It has been suggested that the *Arabidopsis* genome experienced three duplication events within the past 250 million years,<sup>54</sup> while the rice genome is believed to have experienced a genome-wide duplication approximately 70 million years ago.<sup>55,56</sup> To investigate the relationship between the *RALF* genes and potential genomic duplications within the genome, the location of the genes in previously identified *Arabidopsis* and rice chromosomal duplications<sup>57,58</sup> was noted. The distributions of the *RALF* genes relative to the corresponding duplicated genomic blocks were also illustrated in *Arabidopsis* (Fig. S3) and rice (Fig. S4). This result suggested that the generation of 17 (50.0% of 34) *Arabidopsis* and 7 (43.7% of 16) rice *RALF* genes could be due to tandem duplication. In *Arabidopsis*, the largest *RALF* gene cluster was located on chromosome 2 and contained four tandemly arrayed members: ie, *At2g19020*, *At2g19030*, *At2g19040* and *At2g19045* (Fig. 2). Phylogenetically, these four genes formed a single sub-clade in Group II, suggesting that they may result from recent tandem duplications. Because Group II also contains genes from other locations (*At3g25165* and *At3g25170* are located on chromosome 3, whereas *At4g13075* is on chromosome 4), these genes may be the result of more ancient duplication events.

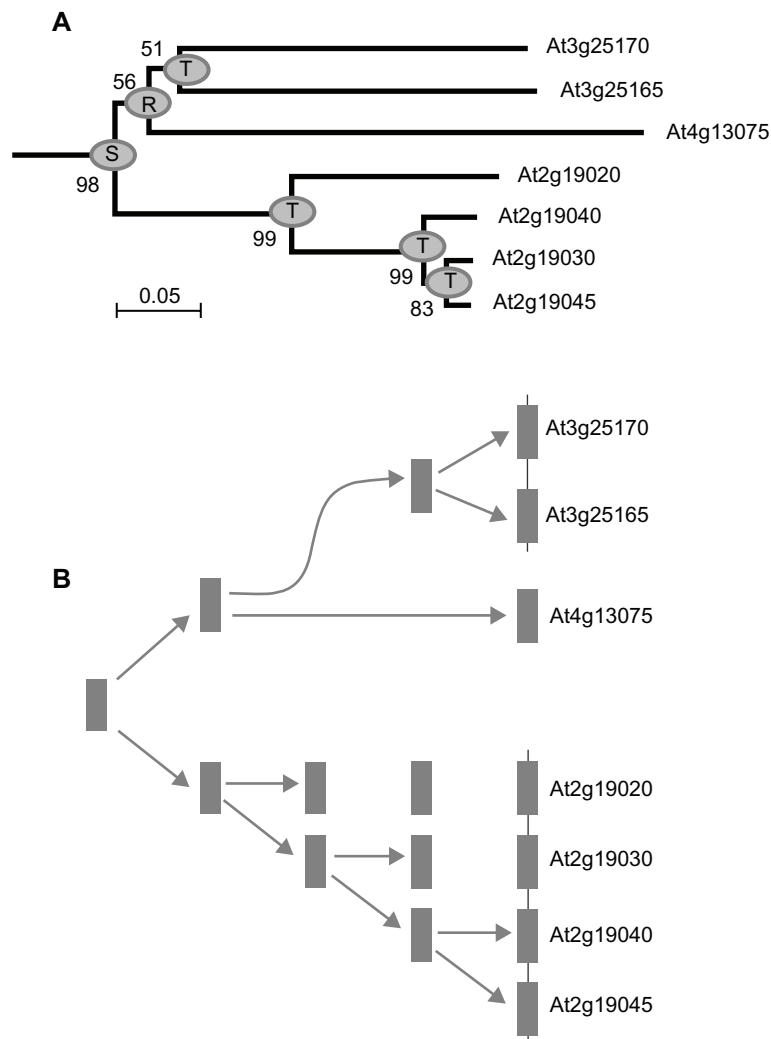
While segmental duplications were not the major factors that led to the expansion of the *RALF* gene family, it might be that dynamic changes occurred following segmental duplication, leading to loss of many of the genes. In contrast to *Arabidopsis* and rice, where 50.0% and 43.7%, respectively, of the *RALFs* were arranged in tandem repeats as described above, considerably fewer *RALFs* were arranged in tandem repeats in poplar (22.7%) and maize (10.5%), indicating that, in these species, *RALFs* mainly emerged by mechanisms other than tandem duplication.

Next, we also investigated the distributions of the unstable and stable *RALFs* in *Arabidopsis*. This result indicated that unstable genes are strongly clustered (about 66.7%), while stable and superstable genes are evenly scattered (or only 38.1% genes clustered) over



**Figure 1.** Evolutionary change in the number of *RALF* genes in *Arabidopsis*, poplar, rice and maize.

**Notes:** The numbers in squares and ellipses represent the maximum numbers of genes in ancestral and extant species (At, *Arabidopsis*; Pt, poplar; Os, rice; Zm, maize), respectively. The numbers with plus and minus indicate the gene gains and losses, respectively, for each branch. N0, lower land plant ancestor; N1, angiosperm ancestor; N2, eudicot ancestor; N3, monocot ancestor.



**Figure 2.** Evolution of the one subgroup of *Arabidopsis* *RALF* genes. **(A)** Phylogenetic relationships. **(B)** Hypothetical origins of seven *Arabidopsis* *RALF* genes by tandem duplication, segmental duplication and retroposition.

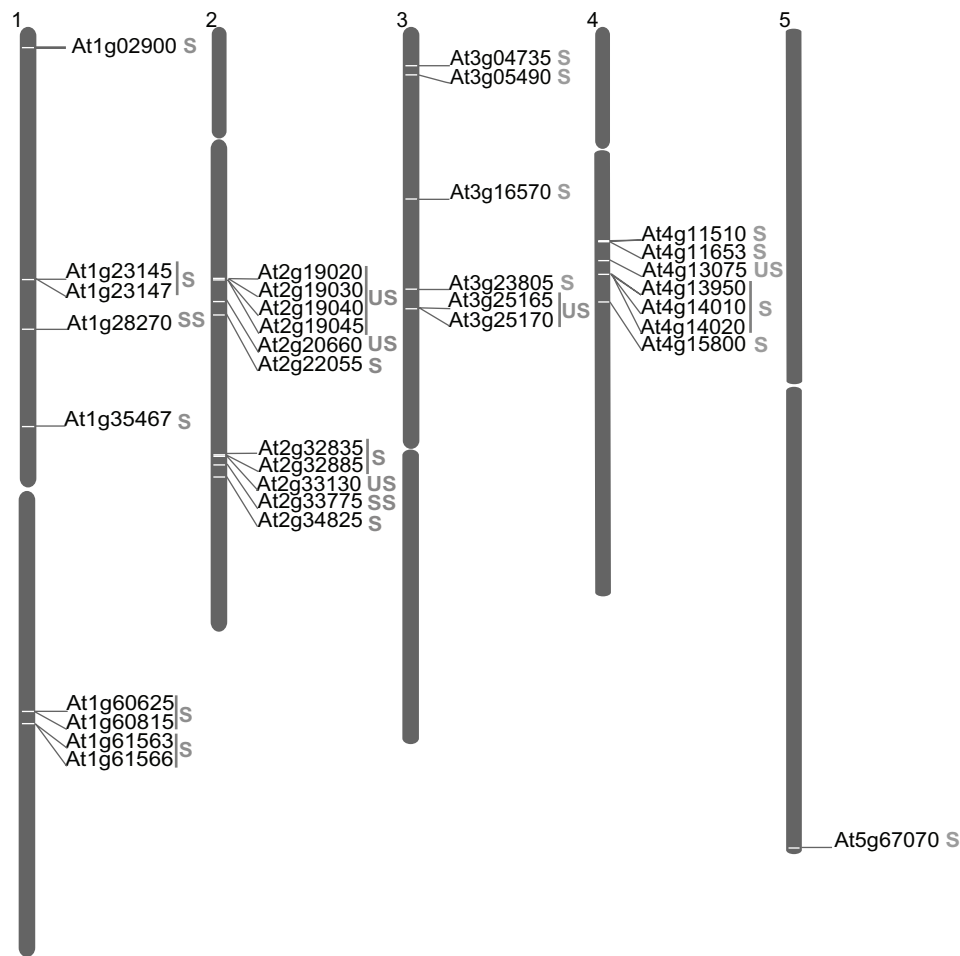
**Notes:** The letters S, R and T on the nodes of the phylogenetic tree indicate the positions where segmental duplication, retroposition and tandem duplication have occurred, respectively.

the chromosomes (Fig. 3). It is clear that the majority of unstable genes in *Arabidopsis* emerged after the most recent whole genome duplication event.<sup>57,59</sup> We also found that, with the exception of *Arabidopsis*, three other species did not contain unstable genes, indicating divergent expansion of the *RALF* genes in different higher plants. In summary, our results suggested that after stable evolution of the *RALF* gene family in *Angiosperms* that followed the divergence from *Tracheophyta* (such as, only two *RALF* genes are identified in *Selaginella moellendorffii*), dramatic expansion had been largely occurred.

In addition, when distantly related species compared, the newly added genes tended to form species-specific clusters or sub-clusters in the eudicots.

For example, seven *Arabidopsis* *RALF* genes formed the most basal cluster within Group II. In Group IV, four poplar *RALF* genes also clustered. This suggested that, as *F-box* genes,<sup>53</sup> the *RALF* genes in different species might have been derived from a series of gene duplication events that occurred after the split of the different lineages. A similar situation was found in the well supported clade of the monocot genes, in which most of the maize and rice genes also formed species-specific clades (such as Groups III and VII, see Fig. S1).

The phylogenetic tree topology revealed several pairs of *RALF* members with a high degree of homology in the terminal nodes of each group, suggesting that they were putative paralogous pairs (Fig. S1).



**Figure 3.** Chromosomal locations of *Arabidopsis* RALF genes.

**Notes:** Approximate positions of RALFs are displayed on the respective chromosome. Letters denote evolutionary classification. s, stable; ss, superstable; u, unstable.

Totally, 13, 7, 6 and 3 pairs of putative paralogous RALF proteins were identified, accounting for more than 78.8%, 60.9%, 63.2% and 37.5% of the entire family in *Arabidopsis*, poplar, maize and rice, respectively, with sequence identities ranging from 30% to 100% (Table S2). These pairs of RALF members are evolutionarily very closely related, and each pair of genes has very similar structure (Fig. S1), indicating that they originated from duplications. About 38.4% of the paralogous RALF pairs in *Arabidopsis* have very consistent  $K_s$  values (from 0.66038 to 0.74663), suggesting that the duplication events in this species occurred within the last 22.01 to 24.89 million years. This period was consistent with the time when a recent large-scale genome duplication event was thought to have occurred in *Arabidopsis*.<sup>42,60</sup> We also found that duplication of three of six RALF pairs originated from the recent large-scale duplication

events (about 15.4 million years ago) in maize.<sup>40</sup> This suggested that, as plant *Sm* and *OPT* genes,<sup>61,62</sup> the recent genome wide duplication events contributed partially to expansion of the RALFs. In addition, in evolutionary terms, some of these RALF gene duplications appeared to have occurred relatively recently, such as *Poptrdraft673738-Poptrdraft672089* (about 1.53 million years ago) and *Poptrdraft578381-Poptrdraft578382* (about 0.8 million years ago). It might be associated with novel functional divergence and adaptation.

Since codon bias can provides some examples of weak selection at the molecular level. Moreover, several researches have verified that selection on synonymous sites is correlated with stability of mRNA secondary structure, translation efficiency and accuracy, ribosome traffic and protein folding.<sup>63–65</sup> We also verified the codon usage bias of RALF genes.

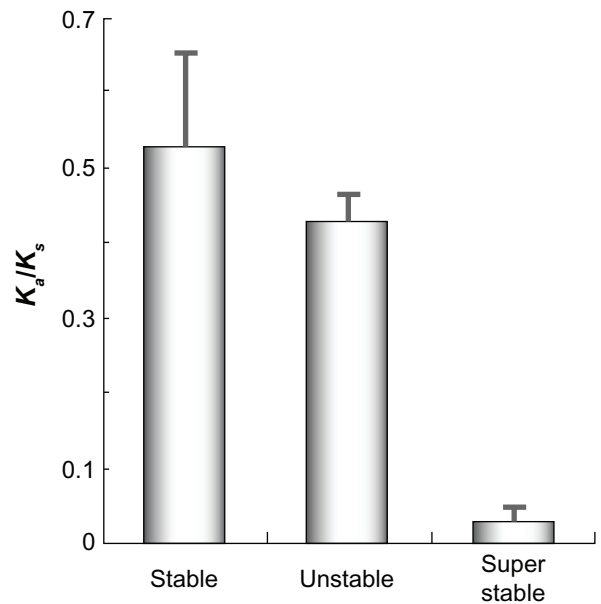
Some information is list in Table S3. In which, CBI and ENC were calculated to measure the degree of codon bias. We can see that CBI showed a marked negative correlation with CBI, so, in this study, ENC was used to measure the degree of codon bias. To determine the relative effects of mutation pressure versus natural selection on codon composition, the relationship between GC3 content and GC1,2 content was examined. The result showed a tendency of positive correlation between GC3 and GC1,2, suggesting that the GC content is most likely the result of mutation pressure since natural selection acts differently on different codon position. In addition, we also confirmed that  $K_s$  was positively correlated with  $K_a$  ( $R^2 = 0.655$ ,  $P < 0.001$ ), and very weakly negatively correlated with ENC and GC3 (but this was not significant) (Fig. S5), implying that codon bias might be a factor in  $K_s$  variation among *RALF* genes and might be under natural selection.

### Different signatures of selection in *RALFs*

To examine whether *RALFs* confer adaptational properties, we determined  $K_a/K_s$  ratios for superstable, stable and unstable genes of *A. thaliana* with *A. lyrata* (Fig. S6).  $K_a/K_s$  ratios of 0.0269 for superstable *RALFs* (Fig. 4) strongly indicated purifying selective pressures. In contrast to that, unstable and stable genes seemed to be closer to neutral selection, as inferred by significantly higher  $K_a/K_s$  ratios (0.5257 and 0.4237) for stable genes and unstable genes, respectively (Fig. 4). We also analyzed the selection properties of the *RALFs* in different Groups. The results showed that the  $K_a/K_s$  ratios of the sequences from the different Groups were significantly different (Table S4). However, despite the differences in  $K_a/K_s$  values, all the estimated  $K_a/K_s$  values were substantially lower than 1, suggesting that the *RALF* sequences within each of the Groups were under strong purifying selection pressure and that positive selection might have acted on only a few sites during the evolutionary process.

### Different expression profiles of the *RALFs* in *Arabidopsis*

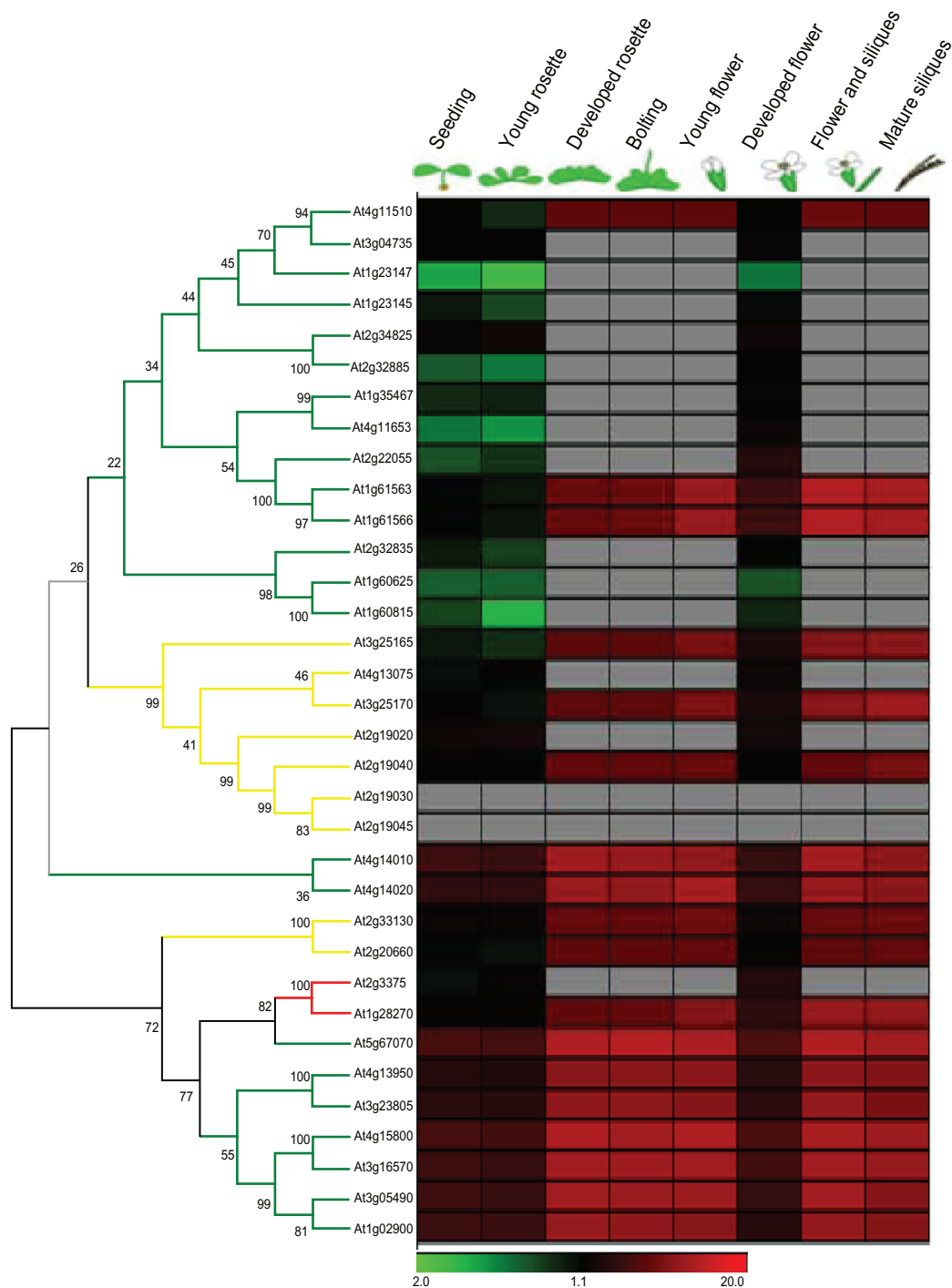
We also examined the expression patterns of the *Arabidopsis* *RALF* genes. The results indicated that



**Figure 4.** Divergence levels of *RALFs* (*A. thaliana* versus *A. lyrata*). **Notes:** Mean  $K_a/K_s$  ratios of stable ( $n = 15$ ), unstable ( $n = 4$ ) and superstable ( $n = 2$ ) are shown. *A. thaliana* genes with a single *A. lyrata* ortholog are included in Figure S6.

the divergent expression profiles were present in stable and unstable *RALFs* across the eight tissues/developmental stages assessed. Furthermore, the stable genes in different evolutionary branches also displayed different expression patterns (Fig. 5). Whether do duplicated genes have similar expression patterns? To answer this question, we investigated their expression profiles and found that none of the pairs of genes shared similar expression patterns (Fig. 5), indicating that substantial neofunctionalization might have occurred during subsequent evolution of the *RALF* duplicated genes. It seems that the expression patterns of the paralogs have diverged during long-term evolution, suggesting functional diversification of the duplicated genes.<sup>66</sup> Such a process ensures the duplicated genes to increase adaptability to environmental changes, thus conferring a possible evolutionary advantage.<sup>67</sup> We also found that over 82% of the assessed genes were likely to be localized in the extracellular space. At2g32885, At2g19030, At2g19040, At1g61563, At1g61566 and At2g19045 have 100% probability of being localized to the extracellular space. For all the other *RALFs*, although the extracellular space was predicted as the most likely location, it was also possible that they were localized to the membranes of organelles such as the cytosol,





**Figure 5.** Expression profiles of the *Arabidopsis* RALF genes.

**Notes:** The dynamic expression profiles were extracted from Genevestigator.<sup>46,47</sup> Green, yellow and red evolutionary branches denote stable, unstable and superstable RALFs in *Arabidopsis*, respectively.

vacuolar membrane or chloroplast. Taken together, while the selected RALFs showed similar subcellular localizations, they differed considerably in their expression profiles, indicating that possible functional diversification may be achieved by selection.

## Conclusion

This study explored the evolutionary process of RALF genes by phylogenetic trees, evolutionary patterns, molecular evolutionary rates, different signatures of selection and the expression patterns of RALFs.



Tandem duplication rather than segmental duplication played a dominant role in the expansion of the *RALF* gene family. *RALFs* were under purifying selection. As well as on the diverse expression patterns of *RALF* genes and predominant extracellular space localization features of *RALF* proteins shed light on several key differences in *RALF* gene family evolution among the four plant species and highlighted the molecular evolution of the *RALF* gene family. All of these may provide a scaffold for future functional analysis of this family.

## Author Contributions

Conceived and designed the experiments: JC. Analysed the data: FS, JC. Wrote the first draft of the manuscript: JC. Contributed to the writing of the manuscript: FS. Agree with manuscript: FS. Jointly developed the structure and arguments for the paper: JC. Made critical revisions and approved final version: JC, FS. All authors reviewed and approved of the final manuscript.

## Funding

This project is supported by grants from the National Science Foundation of China (No. 31100923), the National Science Foundation of Jiangsu Province (BK2011467) and Jiangsu University Senior Personnel Research Grants (10JDG027) to JC.

## Competing Interests

Authors disclose no potential conflicts of interest.

## Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section.

The external blind peer reviewers report no conflicts of interest.

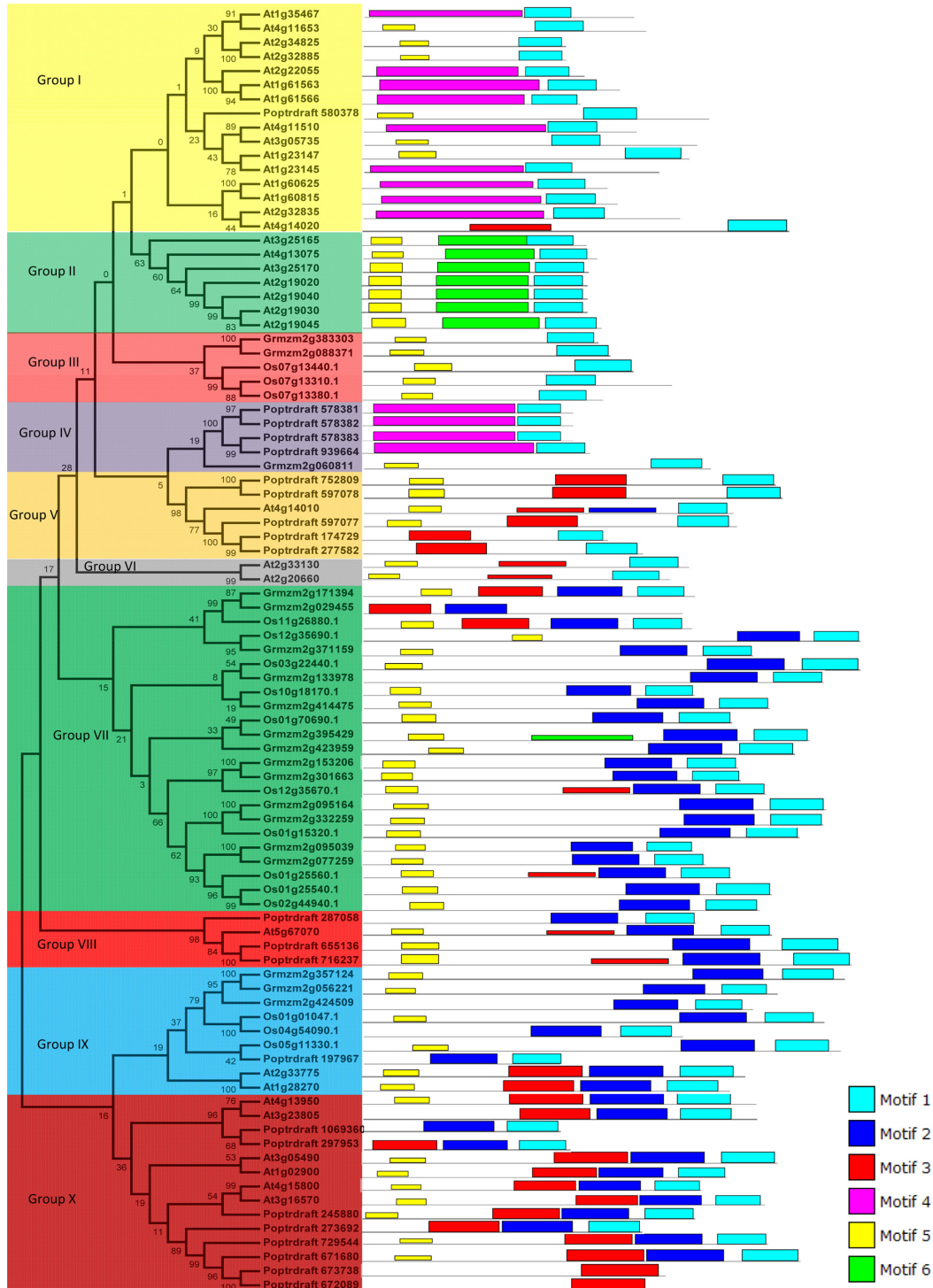
## References

- Matsubayashi Y. *Plant Cell Physiol.* 2011;52:5–13.
- Fletcher JC, Brand U, Running MP, Simon R, Meyerowitz EM. *Science.* 1999;283:1911–4.
- Ryan CA, Pearce G. *Proc Natl Acad Sci U S A.* 2003;100:14577–80.
- Gordon AJ, Minchin FR, James CL, Komina O. *Plant Physiol.* 1999;120:867–78.
- Lay FT, Anderson MA. *Curr Protein Pept Sci.* 2005;6:85–101.
- Yang H, Matsubayashi Y, Nakamura K, Sakagami Y. *Proc Natl Acad Sci U S A.* 1999;96:13560–6549.
- Matsubayashi Y, Sakagami Y. *J Biol Chem.* 2000;275:15520–5.
- Schopfer CR, Nasrallah ME, Nasrallah JB. *Science.* 1999;286:1697–700.
- Takayama S, Shiba H, Iwano M, et al. *Proc Natl Acad Sci U S A.* 2000;97:1920–5.
- Topping JF, Lindsey K. *Plant Cell.* 1997;9:1713–25.
- Narita NN, Moore S, Horiguchi G, et al. *Plant J.* 2004;38:699–713.
- Wen J, Lease KA, Walker JC. *Plant J.* 2004;37:668–77.
- Butenko MA, Patterson SE, Grini PE, et al. *Plant Cell.* 2003;15:2296–307.
- Okuda S, Tsutsui H, Shiina K, et al. *Nature.* 2009;458:357–61.
- Matsuzaki Y, Ogawa-Ohnishi M, Mori A, Matsubayashi Y. *Science.* 2010;329:1065–7.
- Ito Y, Nakanomyo I, Motose H, et al. *Science.* 2006;313:842–5.
- Bedinger PA, Pearce G, Covey PA. *Plant Signal Behav.* 2010;5:1342–6.
- Pearce G, Moura DS, Stratmann J, Ryan CA Jr. *Proc Natl Acad Sci U S A.* 2001;98:12843–7.
- Haruta M, Constabel CP. *Plant Physiol.* 2003;131:814–23.
- Germain H, Chevalier E, Caron S, Matton DP. *Planta.* 2005;220:447–54.
- Wu J, Kurten EL, Monshausen G, Hummel GM, Gilroy S, Baldwin IT. *Plant J.* 2007;52:877–90.
- Combi JP, Kuster H, Journet EP, Hohnjec N, Gamas P, Niebel A. *Mol Plant Microbe Interact.* 2008;21:1118–27.
- Haruta M, Monshausen G, Gilroy S, Sussman MR. *Biochemistry.* 2008;47:6311–21.
- Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB. *BMC Plant Biol.* 2009;9:25.
- Mingossi FB, Matos JL, Rizzato AP, et al. *Plant Mol Biol.* 2010;73:271–81.
- Covey PA, Subbaiah CC, Parsons RL, et al. *Plant Physiol.* 2010;153:703–15.
- Zhang GY, Wu J, Wang XW. *Mol Biol Rep.* 2010;37:3273–81.
- Pearce G, Moura DS, Stratmann J, Ryan CA. *Nature.* 2001;411:817–20.
- Matos JL, Fiori CS, Silva-Filho MC, Moura DS. *FEBS Lett.* 2008;582:3343–7.
- Srivastava R, Liu JX, Guo H, Yin Y, Howell SH. *Plant J.* 2009;59:930–9.
- Escobar NM, Haupt S, Thow G, Boevink P, Chapman S, Oparika K. *Plant Cell.* 2003;15:1507–23.
- Olsen AN, Mundy J, Skriver K. *In Silico Biol.* 2002;2:441–51.
- Bailey TL, Boden M, Buske FA, et al. *Nucleic Acids Res.* 2009;37:W202–8.
- Papadopoulos JS, Agarwala R. *Bioinformatics.* 2007;23:1073–9.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. *Mol Biol Evol.* 2011;28:2731–9.
- Chen K, Durand D, Farach-Colton M. *J Comput Biol.* 2000;7:429–47.
- Cameron JM. *Bioinformatics.* 1999;15:763–4.
- Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. *Bioinformatics.* 2005;21:2101–3.
- Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res.* 2007;35:W506–11.



40. Blanc G, Wolfe KH. *Plant Cell*. 2004;16:1667–78.
41. Yu J, Wang J, Lin W, et al. *PLoS Biol*. 2005;3:e38.
42. Lynch M, Conery JS. *Science*. 2000;290:1151–5.
43. Librado P, Rozas J. *Bioinformatics*. 2009;25:1451–2.
44. Wright F. *Gene*. 1990;87:23–9.
45. Morton BR. *J Mol Evol*. 1993;37:273–80.
46. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. *Plant Physiol*. 2004;136:2621–32.
47. Hruz T, Laule O, Szabo G, et al. *Adv Bioinformatics*. 2008;2008:420747.
48. Sturn A, Quackenbush J, Trajanoski Z. *Bioinformatics*. 2002;18:207–8.
49. Horton P, Park KJ, Obayashi T, et al. *Nucleic Acids Res*. 2007;35:W585–7.
50. Duvick J, Fu A, Muppirala U, et al. *Nucleic Acids Res*. 2008;36:D959–65.
51. Marchler-Bauer A, Anderson JB, Chitsaz F, et al. *Nucleic Acids Res*. 2009;37:D205–10.
52. Marchler-Bauer A, Lu S, Anderson JB, et al. *Nucleic Acids Res*. 2011;39:D225–9.
53. Xu G, Ma H, Nei M, Kong H. *Proc Natl Acad Sci U S A*. 2009;106:835–40.
54. Arabidopsis Genome Initiative. *Nature*. 2000;408:796–815.
55. Yu J, Hu S, Wang J, et al. *Science*. 2002;296:79–92.
56. Paterson AH, Bowers JE, Chapman BA. *Proc Natl Acad Sci U S A*. 2004;101:9903–8.
57. Blanc G, Hokamp K, Wolfe KH. *Genome Res*. 2003;13:137–44.
58. Guyot R, Keller B. *Genome*. 2004;47:610–4.
59. Bowers JE, Chapman BA, Rong J, Paterson AH. *Nature*. 2003;422:433–8.
60. Lynch M, Conery JS. *J Struct Funct Genomics*. 2003;3:35–44.
61. Cao J, Shi F, Liu X, Jia J, Zeng J, Huang G. *J Biomol Struct Dyn*. 2011;28:535–44.
62. Cao J, Huang J, Yang Y, Hu X. *BMC Genomics*. 2011;12:465.
63. Komar AA, Lesnik T, Reiss C. *FEBS Lett*. 1999;462:387–91.
64. Chamary JV, Hurst LD. *Genome Biol*. 2005;6:R75.
65. Gingold H, Pilpel Y. *Mol Syst Biol*. 2011;7:481.
66. Schumann N, Navarro-Quezada A, Ullrich K, Kuhl C, Quint M. *Plant Physiol*. 2011;155:835–50.
67. Cao J, Shi F. *J Biomol Struct Dyn*. 2012; DOI:10.1080/07391102.2012.682207 (accepted).

## Supplemental Data

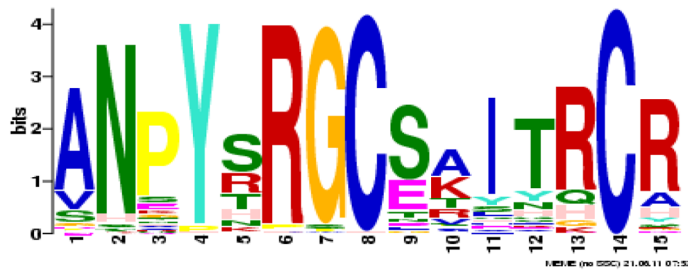


**Figure S1.** Phylogenetic relationships and motif composition of RALF genes in *Arabidopsis*, poplar, rice and maize.

**Notes:** The molecular phylogeny (left panel) was constructed using full length RALF protein sequences from the four species. Numbers associated with branches show bootstrap support values for Neighbor-Joining. The 10 major groups designated from I to X are marked with different colored backgrounds. A schematic representation of conserved motifs (obtained using MEME) in RALF proteins is displayed in the panel on the right. Different motifs are represented by different colored boxes. Details of the individual motifs are in Figure S2.



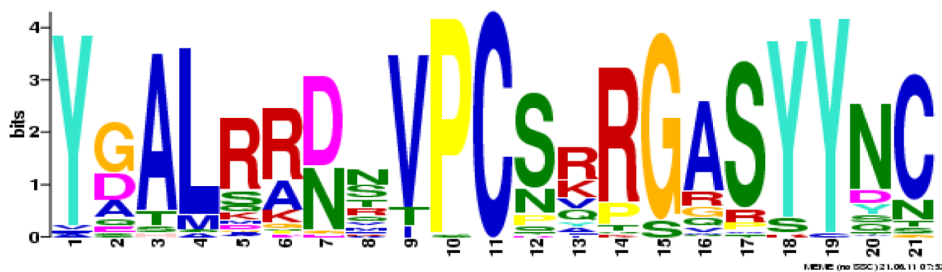
**Motif 1:**



Regular expression:

ANPY[SR]RGC[SE][AK]ITRCR

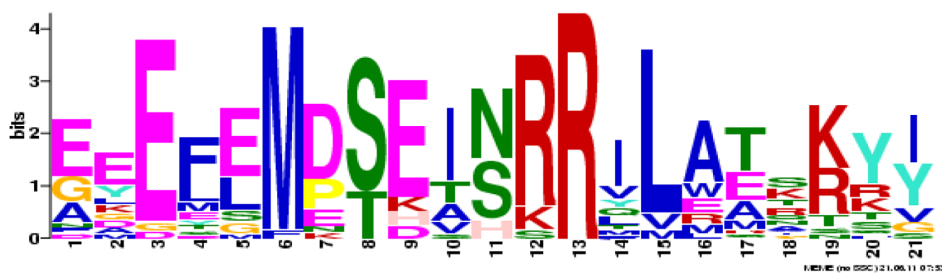
**Motif 2:**



Regular expression:

Y[GD]ALR[RA][DN][NS]VPCSRRGASYINC

**Motif 3:**



Regular expression:

[EG]EEF[EL]M[DP][ST]EI[NS]RRILA[TE]x[KR]Y[IY]

Figure 2. (Continued)

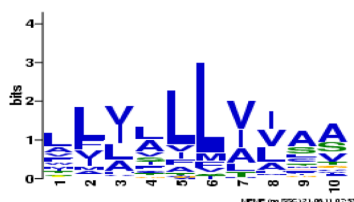
## Motif 4



Regular expression:

[NS][KI][LK][MV][AI]L[GS]L[CA][ML]LM[AF][CL][TA]L[FA]A[GT][NK][AV]EATR[DSY]I  
NYGAIVKGDH[EA]P[FH]C[GD][PK][AKT]HP[CN][VT][KC][TK]

## Motif 5



Regular expression:

LL[IVL][LA]LL[VI][IVL]AA

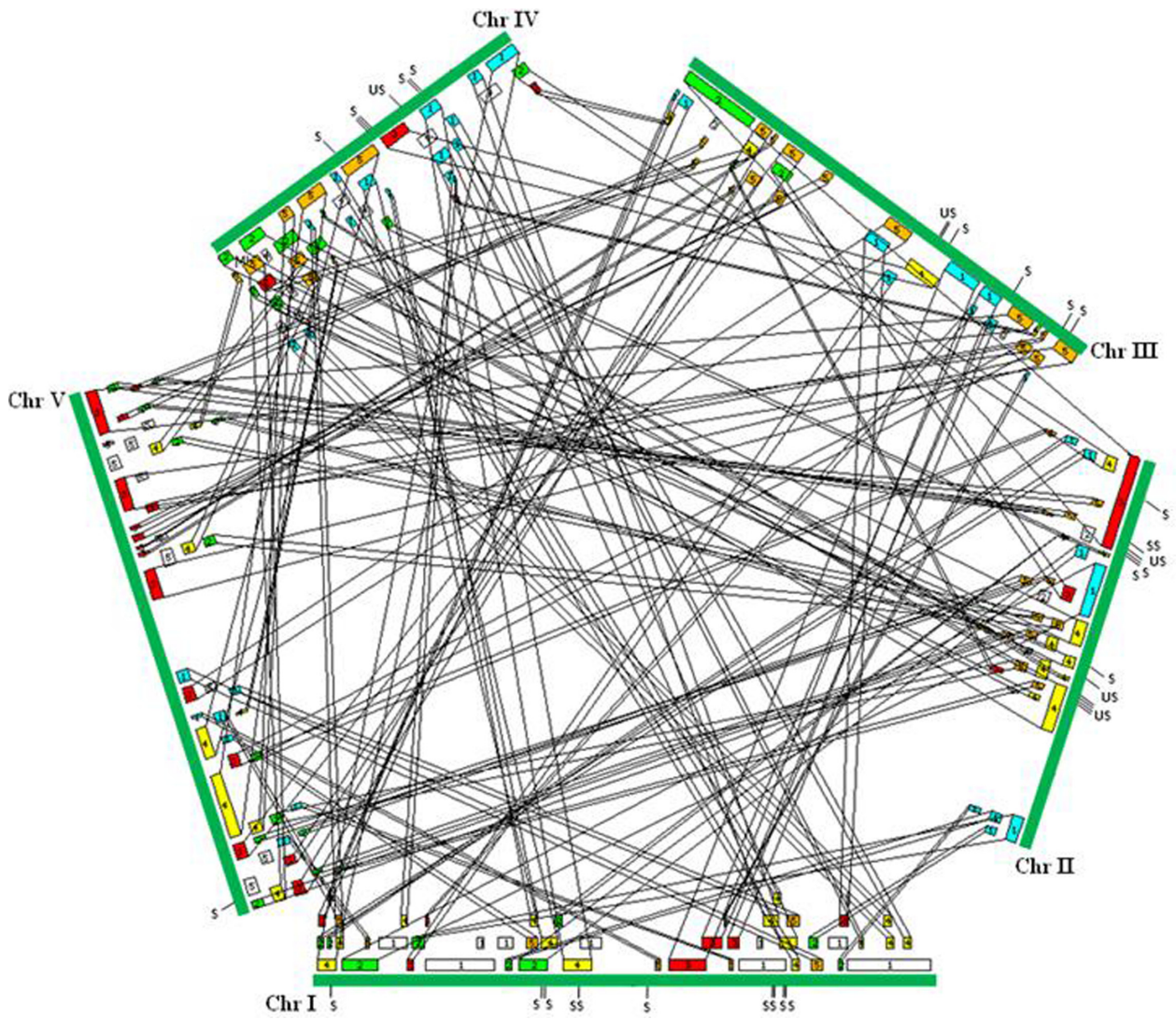
## Motif 6



Regular expression:

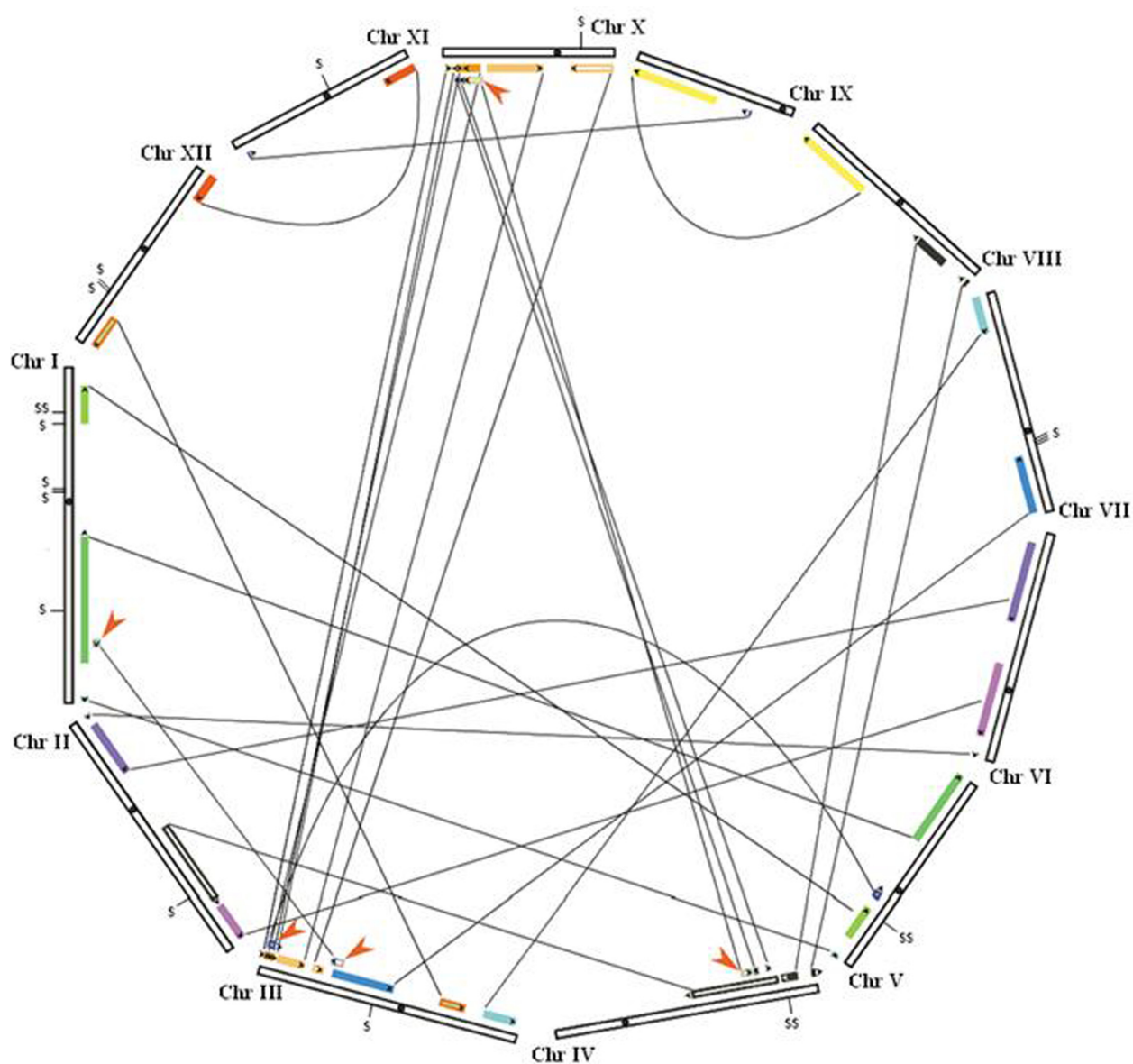
Y[IL][EN][YP]G[AV][IL][DN][KP]C[AL][GR]PNPPPGC[NH]PPG[AS]EQK[NP][PR]

**Figure S2.** Sequence logo and regular expression of the different motifs identified in the *RALF* gene family.



**Figure S3.** Chromosomal locations of the *Arabidopsis* RALF genes.

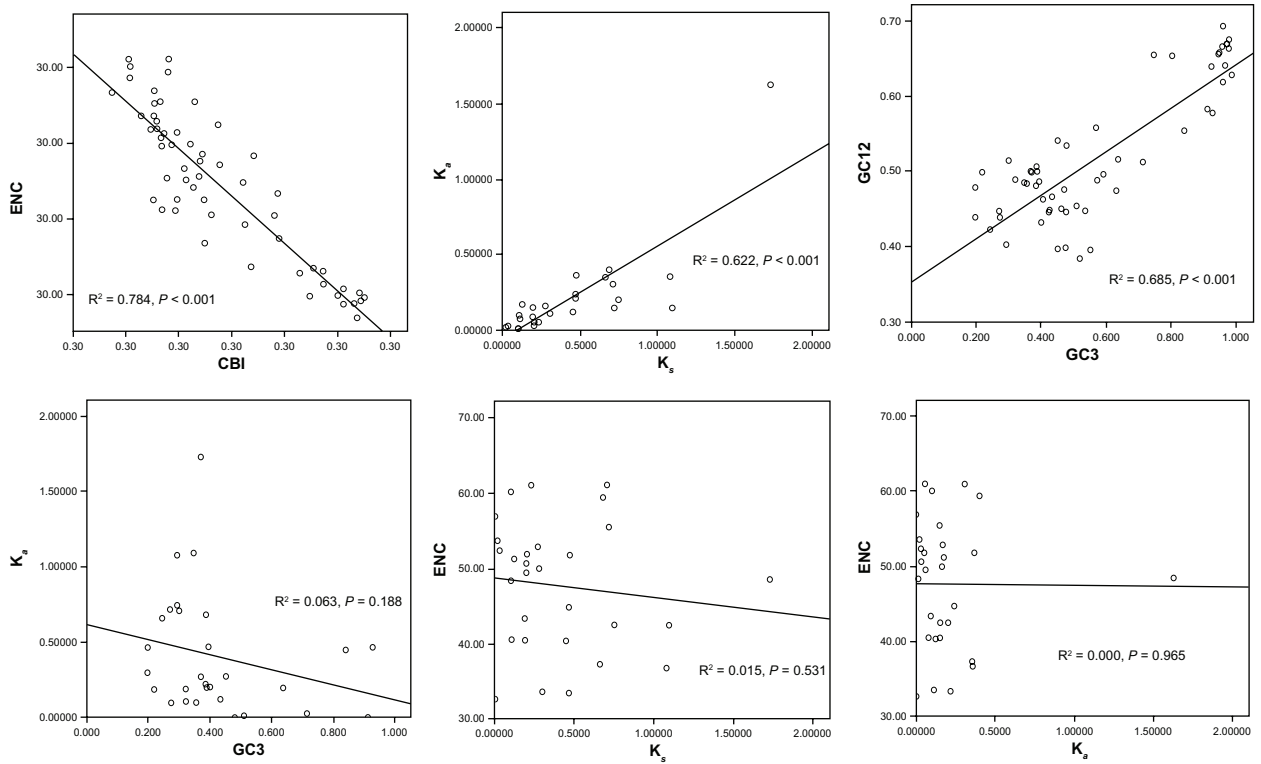
**Notes:** Letters denote evolutionary classification of RALFs. s, stable; ss, superstable; u, unstable. The lines join the segmental duplicated homologous blocks.



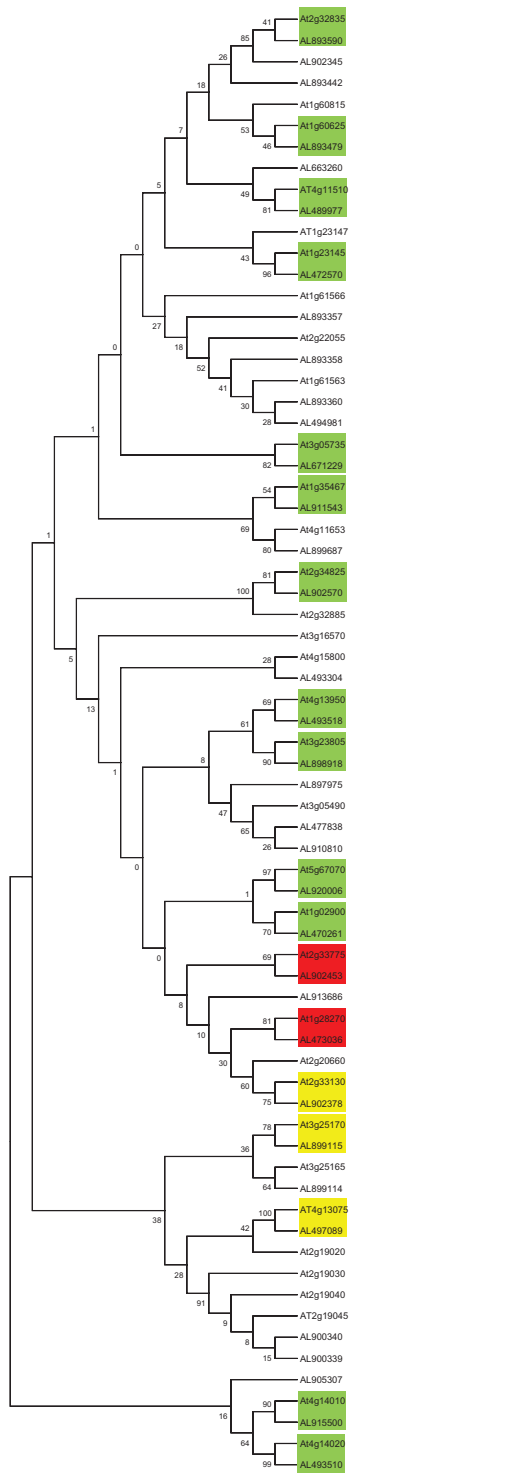
**Figure S4.** Chromosomal locations of the rice *RALF* genes.

**Notes:** Letters denote evolutionary classification of *RALFs*. s, stable; ss, superstable; u, unstable. The lines join the segmental duplicated homologous blocks that are indicated using the same colors.





**Figure S5.** The relationships between ENC and CBI, ENC and  $K_s$ , ENC and  $K_a$ ,  $K_a$  and  $K_s$ , GC1,2 and GC3,  $K_s$  and GC3.



**Figure S6.** NJ tree generated using RALF protein sequences of *A. thaliana* and *A. lyrata*.  
**Note:** Numbers at branches indicate bootstrap values (1000 replicates). Boxed sequences designate proteins used for  $K_a/K_s$  ratios: green = stable, yellow = unstable and red = superstable.

**Table S1.** Number of *RALF* genes in lower plants.

Species	Number of RALFs	Gene ID*
<i>Selaginella moellendorffii</i>	2	9636436, 9661207
<i>Physcomitrella patens</i>	3	5920213, 5945963, 5942388

**Note:** \*GenBank ID.

**Table S2.** Pairwise identities and inference of duplication time in paralogous pairs of RALF genes from *Arabidopsis*, poplar, maize and rice.

Paralogous pairs	Score	Expect	Identities	Positives	Gaps	$K_a$	$K_s$	Date (million years ago)
At1g35467-At4g11653	87.8	2.00E-23	48/90 (53%)	59/90 (66%)	4/90 (4%)	0.35552	0.66038	22.01
At2g34825-At2g32885	122	5.00E-34	58/71 (82%)	64/71 (90%)	0/71 (0%)	0.11508	0.29887	9.96
At1g61563-At1g61566	128	8.00E-36	63/73 (86%)	67/73 (92%)	1/73 (1%)	0.09239	0.18812	6.27
At4g11510-At3g04735	77.4	4.00E-20	48/93 (52%)	57/93 (61%)	14/93 (15%)	0.30588	0.70782	23.59
At1g23147-At1g23145	68.6	2.00E-17	40/82 (49%)	51/82 (62%)	16/82 (20%)	0.39871	0.68397	22.79
At1g60625-At1g60815	120	4.00E-33	55/80 (69%)	66/80 (83%)	0/80 (0%)	0.17395	0.1231	4.1
At2g32835-At4g14020	26.9	6.00E-05	14/46 (30%)	21/46 (46%)	0/46 (0%)	1.62731	1.72676	57.56
At2g19030-At2g19045	115	6.00E-32	70/72 (97%)	71/72 (99%)	0/72 (0%)	0.01142	0.09964	3.32
At2g33130-At2g20660	108	2.00E-29	62/104 (60%)	70/104 (67%)	8/104 (8%)	0.24131	0.46629	15.54
At2g33775-At1g28270	179	6.00E-51	85/107 (79%)	94/107 (88%)	0/107 (0%)	0.15015	1.09122	36.37
At4g13950-At3g23805	146	1.00E-40	79/118 (67%)	90/118 (76%)	5/118 (4%)	0.20268	0.74663	24.89
At3g05490-At1g02900	129	1.00E-35	69/120 (58%)	83/120 (69%)	3/120 (3%)	0.3585	1.07798	35.93
At4g15800-At3g16570	153	9.00E-43	88/137 (64%)	98/137 (72%)	21/137 (15%)	0.149	0.71762	23.92
Poptrdraft578381-Poptrdraft578382	155	5.00E-44	73/75 (97%)	73/75 (97%)	0/75 (0%)	0.01865	0.01465	0.8
Poptrdraft578383-Poptrdraft939664	151	1.00E-42	72/72 (100%)	72/72 (100%)	0/72 (0%)	0	0	0
Poptrdraft52809-Poptrdraft597078	146	9.00E-41	82/124 (66%)	94/124 (76%)	4/124 (3%)	0.16283	0.27758	15.25
Poptrdraft174729-Poptrdraft277582	170	3.00E-48	84/84 (100%)	84/84 (100%)	0/84 (0%)	0	0	0
Poptrdraft655136-Poptrdraft716237	212	2.00E-60	104/135 (77%)	111/135 (82%)	5/135 (4%)	0.12351	0.45034	24.74
Poptrdraft1069360-Poptrdraft297953	113	2.00E-31	52/59 (88%)	57/59 (99%)	0/59 (0%)	0.05688	0.1977	10.86
Poptrdraft673738-Poptrdraft672089	162	5.00E-46	78/83 (94%)	79/83 (95%)	0/83 (0%)	0.02725	0.02791	1.53
Gmzm2g383303-Gmzm2_g088371	115	6.00E-32	55/72 (76%)	62/72 (86%)	0/72 (0%)	0.16484	0.26952	20.73
Gmzm2g171394-Gmzm2g029455	89.7	8.00E-24	44/49 (90%)	46/49 (94%)	2/49 (4%)	0.36496	0.47344	36.42
Gmzm2g153206-Gmzm2g301663	134	4.00E-37	80/102 (78%)	81/102 (79%)	11/102 (11%)	0.05501	0.22857	17.58
Gmzm2g095039-Gmzm2g077259	164	4.00E-46	94/114 (82%)	96/114 (84%)	7/114 (6%)	0.03113	0.19982	15.37
Gmzm2g095164-Gmzm2g332259	175	2.00E-49	102/121 (84%)	104/121 (86%)	5/121 (4%)	0.0507	0.20262	15.58
Gmzm2g357124-Gmzm2g056221	164	5.00E-46	94/120 (78%)	103/120 (86%)	3/120 (3%)	0.09789	0.10101	7.77
Os07g13310.1-Os07g13380.1	118	1.00E-32	55/76 (72%)	65/76 (86%)	0/76 (0%)	0.15137	0.19107	14.69
Os01g25540.1-Os02g44940.1	112	2.00E-30	77/90 (86%)	80/90 (89%)	2/90 (2%)	0.07853	0.10795	8.3
Os01g010147.1-Os04g54090.1	87.8	4.00E-23	61/113 (54%)	63/113 (56%)	18/113 (16%)	0.21539	0.46804	36

Table S3. Codons information for the paralogous pairs of *RALF* genes list in Table S2.

Sequence_name	ENC	CBI	SChi2	GC1,2	GC3 s	GC3	GC
At1g35467	37.336	0.689	0.888	0.423	0.212	0.244	0.363
At4g11653	33.567	0.636	0.979	0.439	0.182	0.198	0.359
At2g34825	43.384	0.687	0.828	0.499	0.152	0.219	0.406
At2g32885	61.000	0.480	0.682	0.514	0.250	0.301	0.443
At1g61563	59.383	0.480	0.551	0.506	0.388	0.386	0.466
At1g61566	51.270	0.497	0.716	0.467	0.411	0.434	0.456
At4g11510	48.505	0.543	0.660	0.499	0.329	0.372	0.457
At3g04735	48.336	0.641	0.775	0.439	0.232	0.274	0.384
At1g23147	44.772	0.622	0.796	0.479	0.205	0.198	0.385
At1g23145	42.469	0.547	0.687	0.485	0.293	0.347	0.439
At1g60625	42.506	0.452	0.632	0.403	0.269	0.293	0.366
At1g60815	36.693	0.549	0.835	0.403	0.278	0.293	0.366
At2g32835	55.438	0.530	0.605	0.448	0.242	0.271	0.389
At4g14020	53.626	0.429	0.494	0.455	0.481	0.509	0.473
At2g19030	56.905	0.453	0.567	0.535	0.486	0.479	0.516
At2g19045	49.911	0.522	0.647	0.541	0.486	0.452	0.511
Grzm2g383303	32.689	0.728	1.100	0.583	0.921	0.911	0.692
Grzm2g088371	40.404	0.680	1.002	0.554	0.847	0.840	0.649
Os07g13310	49.507	0.468	0.526	0.516	0.617	0.636	0.556
Os07g13380	52.350	0.575	0.644	0.512	0.714	0.713	0.579
Poptdraft578381	52.878	0.458	0.589	0.500	0.333	0.368	0.456
Poptdraft578382	51.817	0.446	0.643	0.487	0.348	0.395	0.456
Poptdraft578383	61.000	0.405	0.480	0.481	0.370	0.385	0.449
Poptdraft939664	50.685	0.466	0.599	0.500	0.353	0.389	0.463
Poptdraft752809	51.849	0.459	0.523	0.432	0.378	0.400	0.421
Poptdraft597078	60.079	0.407	0.443	0.484	0.330	0.355	0.441
Poptdraft174729	40.482	0.561	0.795	0.489	0.313	0.321	0.433
Poptdraft277582	40.482	0.561	0.795	0.489	0.313	0.321	0.433
Grzm2g171394	33.424	0.754	1.184	0.578	0.923	0.927	0.694
Grzm2g029455	41.020	0.493	0.629	0.655	0.716	0.748	0.686
Grzm2g153206	29.812	0.747	1.270	0.639	0.932	0.924	0.734
Grzm2g301663	26.809	0.837	1.545	0.664	0.984	0.976	0.768
Grzm2g095164	30.170	0.841	1.494	0.675	0.985	0.978	0.776
Grzm2g332259	29.469	0.849	1.559	0.669	0.978	0.971	0.770
Grzm2g095039	31.342	0.773	1.323	0.658	0.955	0.947	0.754
Grzm2g077259	28.643	0.831	1.455	0.656	0.953	0.945	0.752
Os01g25540	30.591	0.810	1.438	0.666	0.957	0.957	0.763
Os02g44940	29.803	0.801	1.332	0.693	0.966	0.958	0.781
Poptdraft655136	51.192	0.471	0.520	0.446	0.400	0.424	0.439
Poptdraft716237	45.475	0.537	0.617	0.463	0.375	0.407	0.444
Grzm2g357124	28.604	0.810	1.340	0.641	0.972	0.966	0.749
Grzm2g056221	29.062	0.842	1.385	0.629	0.993	0.986	0.748
Os01g10470	33.092	0.773	1.098	0.619	0.964	0.959	0.732
Os04g54090	39.185	0.625	0.883	0.654	0.806	0.804	0.704





At2g33775	47.177	0.577	0.622	0.474	0.626	0.631	0.526
At1g28270	45.116	0.513	0.629	0.446	0.462	0.477	0.456
At4g13950	53.593	0.452	0.500	0.448	0.519	0.535	0.477
At3g23805	49.805	0.487	0.570	0.450	0.427	0.462	0.454
Poptrdraft297953	41.193	0.469	0.793	0.449	0.400	0.426	0.441
Poptrdraft1069360	44.057	0.527	0.836	0.396	0.516	0.552	0.448
At3g05490	42.556	0.496	0.575	0.496	0.569	0.592	0.528
At1g02900	56.627	0.373	0.378	0.476	0.453	0.471	0.474
At4g15800	58.532	0.407	0.409	0.488	0.561	0.573	0.516
At3g16570	55.177	0.454	0.446	0.558	0.567	0.568	0.561
Poptrdraft673738	47.574	0.540	0.689	0.399	0.434	0.476	0.425
Poptrdraft672089	45.379	0.477	0.589	0.399	0.442	0.476	0.425
At2g33130	55.418	0.465	0.534	0.384	0.396	0.519	0.429
At2g20660	46.571	0.510	0.586	0.397	0.418	0.451	0.415
Mean ± SE	44.410 ± 9.869	0.578 ± 0.139	0.805 ± 0.329	0.514 ± 0.088	0.541 ± 0.267	0.558 ± 0.253	0.529 ± 0.137

**Table S4.** Likelihood values and parameter estimates for the *RALF* genes.

Gene branches	Model	$K_a/K_s$	Log-likelihood	Positive selection sites
Group I	M8	0.5206	-4162.59	Not found
	M8a	0.4827	-4163.25	Not found
	M7	0.5138	-4162.07	Not found
	M5	0.5723	-4166.45	35,66,69,77,79,90,106
	MEC	0.6459	-4096.26	39,43,66,69,77,79,
Group II	M8	0.6145	-1122.75	7,13,14,19,20,23,24,25,38,49,50,52,53,57,70,73,76
	M8a	0.4435	-1123.65	Not found
	M7	0.4398	-1123.63	Not found
	M5	0.4846	-1124.25	7,52,57,76
	MEC	0.7122	-1115.32	4,7,13,14,20,23,24,25,30,38,49,52,53,57,61,73,76
Group III	M8	0.6031	-1137.14	4,27,28,32,34,51,56,64,72,74,78
	M8a	0.4139	-1136.04	Not found
	M7	0.4483	-1136.32	Not found
	M5	0.4737	-1138	64,78
	MEC	0.6697	-1131.68	4,10,25,27,28,32,34,36,51,54,55,56,58,64,72,74,77,78
Group IV	M8	0.4432	-802.603	Not found
	M8a	0.4913	-803.102	Not found
	M7	0.4271	-802.584	Not found
	M5	0.4917	-803.795	Not found
	MEC	0.5384	-797.179	48,58,60,68,72
Group V	M8	0.372	-1587.53	Not found
	M8a	0.4232	-1588.03	Not found
	M7	0.3795	-1587.36	Not found
	M5	0.4154	-1587.73	8,
	MEC	0.6176	-1580.61	3,5,6,8,14,17,19,21,22,24,35,38,43,45,54,60,67,72,76,78,86,88,93,101,111,121
Group VII	M8	0.4258	-7178.9	Not found
	M8a	0.3071	-7200.06	Not found
	M7	-	-	-
	M5	0.3663	-7211.88	Not found
	MEC	0.4086	-7047.72	68,71,75,77,
Group VIII	M8	0.4547	-1471.73	3,14,15,17,18,30,31,36,48,52,55,61,109,112,123
	M8a	0.2244	-1466.05	Not found
	M7	0.2281	-1466.04	Not found
	M5	0.2781	-1467.72	Not found
	MEC	0.4315	-1472.78	16,17,36,52,62,112
Group IX	M8	0.3449	-2778.46	Not found
	M8a	0.342	-2776.4	Not found
	M7	0.3533	-2778.6	Not found
	M5	0.3557	-2779.98	Not found
	MEC	0.4434	-2733.01	25,31,37,38,39,40,43,44,81,84
Group X	M8	0.3125	-3496.94	Not found
	M8a	0.2881	-3500.02	Not found
	M7	0.3155	-3500.78	Not found
	M5	0.3206	-3504.9	Not found
	MEC	0.3524	-3441.27	7,9,10,23,25