Evolution of Plant Ribosome-Inactivating Proteins

Willy J. Peumans and Els J.M. Van Damme

Abstract This contribution presents an updated analysis of the evolution of ribosome-inactivating proteins (RIPs) in plants. All evidence suggests that an ancestor of modern seed plants developed the RIP domain at least 300 million years ago. This ancestral RIP domain gave rise to a direct lineage of type 1 RIPs (i.e., primary type 1 RIPs) still present today in many monocots and at least one dicot. In a later stage, a plant succeeded in fusing the RIP domain to a duplicated ricin-B domain acquired from a bacterium. The resulting ancestral type 2 RIP gave rise to all modern type 2 RIPs and by domain deletion, to different lines of "secondary" type 1 RIPs and ricin-B type lectins. In the recent past, at least three other domain fusions took place in the Poaceae family, whereby type AC1 (type 3), type AC2, and type AD chimeric forms were generated.

1 Introduction

Plant ribosome-inactivating proteins (RIPs) are a fairly extended and heterogeneous family of proteins characterized by the presence of a domain equivalent to the toxic A-chain of ricin (or A-subunit of the bacterial Shiga toxins). Basically, the plant RIPs can be subdivided into holoenzymes and chimero-enzymes. Holoenzymes or type 1 RIPs consist solely of a RIP domain whereas the chimero-enzymes are built up of an N-terminal RIP domain linked (at least in the gene) to an unrelated C-terminal domain. Depending on the nature of the latter chain, the chimeric forms are referred to as type 2 RIPs (with a lectinic B-chain) and type 3 RIPs

W.J. Peumans and E.J.M. Van Damme (✉)

Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
e-mail: ElsJM.VanDamme@UGent.be

(with an unidentified C-terminal domain). Both type 1 and type 2 RIPs are quite common in plants whereas hitherto only a single type 3 RIP has been isolated and characterized, namely the barley JIP60 (Chaudhry et al. 1994). However, recent genome and transcriptome data revealed the occurrence of homologs in some other Poaceae. Moreover, there is also evidence for yet another chimeric form in rice and *Brachypodium distachyon*. Since none of these "putative novel" RIPs has been studied at the biochemical level, it is precocious to introduce a new nomenclature. Therefore, JIP60 and its homologs will be referred to as type AC proteins, and the additional form found in rice and *Brachypodium* as type AD proteins, to emphasize the fact that they possess a different C-terminal domain.

The issue of the molecular evolution of plant RIPs was already discussed in numerous research and review papers (Barbieri et al. 1993; Peumans et al. 2001; Van Damme et al. 2001; Stirpe and Battelli 2006). Though several aspects of the overall evolution are fairly well understood, some important questions remain to be answered especially with respect to the origin of the RIP domain, the relationships between type 1 and type 2 RIPs, and the origin of the type 3 RIP. One of the major problems encountered in the study of the evolution and phylogeny concerns the limited number of sequences and the patchy taxonomic distribution of plant RIPs. Fortunately, the wealth of information provided by genome and transcriptome sequencing programs allows composing a more detailed overview of the occurrence of RIPs in plants and reassessing the interrelationships between the different subgroups. Moreover, the eventual origin of the RIP domain itself as well as the B-chain of type 2 RIPs could also be revised using the sequence information made available for bacteria and eukaryotes other than plants.

This contribution aims to make an updated comprehensive analysis of the overall evolution of plant RIPs. Therefore, an as-complete-as-possible set of sequences was retrieved from the publicly accessible databases and subsequently subjected to a preliminary phylogenetic analysis (using CLUSTALW). Considering the limitations of this method, the results should be interpreted with care. However, the outlines of the analyses give a fairly accurate idea and place the overall evolution of RIPs in a novel perspective. Moreover, the data generated here provide a firm basis for an in-depth phylogenetic analysis with a more performing program.

2   General Overview of the Taxonomic Distribution of A and B Domains within the Viridiplantae

According to the data published in previous research and review papers, the occurrence within the Viridiplantae of both the RIP and the ricin-B domain is confined to the Magnoliophyta (flowering plants) (Van Damme et al. 1998; 2001). To check whether these domains possibly occur in other taxa, a comprehensive analysis of the publicly accessible databases was made. BLAST searches (using

different type 1 and type 2 RIP sequences[1] as queries) in the completed genomes of *Chlamydomonas reinhardtii*, *Chlorella* sp., *Micromonas* sp., *Ostreococcus* sp., and *Volvox carteri* yielded no positive hits indicating, though not proving, that Chlorophyta (green algae) genomes acquired neither the RIP nor the ricin-B domain. Within the Embryophyta, proteins/genes with a RIP domain are apparently confined to the Spermatophyta (seed plants). No A domain could be identified, indeed, in any member of the Anthocerotophyta, Bryophyta, Marchantiophyta, or Euphyllophyta other than Spermatophyta. Though, due to the limited sequence information, one cannot draw definitive conclusions regarding the possible occurrence of this domain in these major taxonomic groups, the apparent absence of the RIP domain in the completed genomes of the moss *Physcomitrella patens* and the club moss *Selaginella moellendorffii* is certainly indicative. Contrary to the A domain, there is compelling evidence of the expression of proteins with a typical ricin-B domain in the liverwort *Marchantia polymorpha*. Analysis of the transcriptome database revealed that thalli and sexual organs of *M. polymorpha* express a set of at least three different proteins comprising two in tandem arrayed ricin-B domains (and hence can be considered the equivalent of the B-chain of a type 2 RIP). One of these expressed proteins has -apart from the N-terminal Met residue- exactly the same sequence as the N-terminus of a galactose-binding lectin isolated from thallus tissue (EVD unpublished results), which leaves no doubt that this liverwort actually expresses a carbohydrate-binding protein of the ricin-B family. It should be noted here that the purified lectin is synthesized without signal peptide and undergoes, apart from the removal of the methionine, no processing at its N-terminus. This implies that the *Marchantia* lectins are unlike all other documented plant lectins of the ricin-B family (which are synthesized with a signal peptide and follow the secretory pathway) (Van Damme et al. 2001). *Marchantia* lectins are synthesized on free ribosomes in the cytoplasm and accordingly destined to reside in the cytoplasmic and/or nuclear compartment.

Hitherto, all purified plant RIPs and cloned plant RIP genes were obtained from Magnoliophyta (flowering plants). No homologs were isolated from or identified in any other seed plant. Transcriptome analyses also yielded no evidence of the expression of RIP genes in Coniferophyta (approximately 800,000 entries), Cycadophyta (22,000 entries), and Ginkgophyta (21,000 entries). In contrast, a recently deposited transcriptome database of *Gnetum gnemon* (10,700 entries) contains a set of three expressed sequence tags (ESTs) encoding two different type 2 RIPs. The latter finding is important because it demonstrates for the first time the occurrence of RIP genes in a seed plant outside the flowering plants. Taking into consideration the very large number of deposited EST sequences (about 800,000 in total) of different *Pinus* and *Picea* species, it seems unlikely that RIP genes are present in the genome of most modern Coniferophyta. Due to the

---

[1]Sequences of all RIPs used in this study can be retrieved from: http://www.molecularbiotechnology.ugent.be/publications/VanDamme2010A/.

relatively small number of entries no such conclusion can be drawn yet for the Cycadophyta (22,000 entries) and Ginkgophyta (21,000 entries).

Comprehensive BLAST searches of plant transcriptome databases yielded several sequences encoding proteins consisting of a single ricin-B domain (i.e., corresponding to one half of the B-chain of type 2 RIPs). At first sight, the identification of these proteins was exciting because it could give valuable hints with regard to the origin of the B-chain of type 2 RIPs. However, a closer examination indicated that these sequences are not encoded by the plant genome but by a contaminating fungus or other eukaryotic symbiont/parasite. For example, a strongly conserved protein expressed in roots of wheat and poplar, and stolons of potato turns out to be 94% identical to a large set (>250) of ESTs present in the transcriptome of *Hartmannella vermiformis* (a protozoan belonging to the Euamoebida). Hence, it is almost certain that the sequences encoding these "root-specific" proteins are derived from a contaminating amoeba.

3 Overview of the Taxonomic Distribution of A and B Domains within the Magnoliophyta (Flowering Plants)

3.1 *"Classical" Type 2 RIPs (AB proteins)*

Hitherto, only a relatively small set of type 2 RIPs (<40) has been purified and characterized. Moreover, since all these proteins were isolated from a rather limited number of seed plants (*Ricinus communis*, *Abrus* sp., *Adenia* sp., *Cinnamomum camphora*, *Sambucus* sp., *Viscum* sp., *Momordica charantia*, *Trichosanthes* sp., *Bryonia dioica*, *Panax ginseng*, *Ximenia americana*, *Iris hollandica*, and *Polygonatum multiflorum*) belonging to only 13 genera, it is generally believed that type 2 RIPs are scarcely distributed among flowering plants. However, taking this into consideration, by analogy to other plant lectins, the currently documented taxonomical distribution might well be underestimated. This is because only highly expressed type 2 RIPs have a reasonable chance of being discovered by routinely applied techniques. To address this issue, we searched the publicly accessible databases (mostly transcriptomes) for the occurrence of type 2 RIPs and related proteins with ricin-B domain(s). Evidence was obtained for the expression of one or more genuine type 2 RIP homologs in several species with no "RIP history." The list of novel species includes flowering plants from most major taxa: Ranunculaceae (*Adonis aestivalis*); Asteraceae (*Helianthus* sp., *Centaurea* sp., *Artemisia annua*), Ericales (Actinidiaceae: *Actinidia deliciosa* or kiwi fruit; Polemoniaceae: *Ipomopsis aggregata*; Theaceae: *Camellia sinensis* or tea plant); Cucurbitaceae (*Cucumis sativus*), Rosaceae (*Malus domestica* or apple); Malvaceae (*Gossypium* sp.), Sapindaceae (*Paullinia cupana* or guarana), Poaceae (*Triticum aestivum*, *Saccharum officinarum*, *Sorghum* sp., *Zea mays*, *Panicum virgatum*), and Arecaceae (*Elaeis guineensis* or oil palm).

3.2 Other Proteins with Ricin-B Domains

Proteins consisting exclusively of polypeptides equivalent to the B-chain of type 2 RIPs have been isolated from several *Sambucus* species (e.g., fruit lectin SNA-IV) (Girbés et al. 2004; Van Damme et al. 1997b) and from cucumber (*C. sativus*) roots (XSP30) (Masuda et al. 1999). Moreover, it seems quite likely that several lectins previously isolated from different *Euphorbia* species are built up of B-chains only (Stirpe et al. 1993). The searching of databases yielded only a few novel B-type proteins. Transcriptome analysis confirmed that *Euphorbia esula* expresses several proteins consisting of B-chains only. In addition, a set of five different but closely related B-chain proteins (most of which are expressed) could be identified in the genome of *Populus trichocarpa*. The same genome also contains a gene encoding a protein consisting of a single ricin-B domain, but no corresponding ESTs could be retrieved.

4 Molecular Evolution of Type 2 RIPs

4.1 General Observations Concerning the Taxonomic Distribution of Type 2 RIPs and the Occurrence of Multiple Paralogs

In spite of the tremendous amount of (multidisciplinary) research devoted to ricin and related proteins, the molecular evolution and especially the evolutionary origin of type 2 RIPs is still far from understood. At present, no direct evolutionary link can be made between type 2 RIPs from modern seed plants and any protein from lower Viridiplantae (or any other organism). It seems likely, therefore, that type 2 RIPs were developed by a direct ancestor of modern seed plants. Until recently type 2 RIPs were believed to occur exclusively in flowering plants (Van Damme et al. 2008). However, novel data leave no doubt that genuine homologs are also expressed in at least one species of the Gnetophyta. On the basis of this updated taxonomic distribution (summarized in Tables 1 and 2) one can reasonably conclude that type 2 RIP gene(s) were already present in the common ancestor of the Gnetophyta and Magnoliophyta, and accordingly exist for over 300 million years (Palmer et al. 2004). The reasonably high sequence identity between the homologs from *G. gnemon* and those from flowering plants can be perfectly explained by a classical vertical inheritance along the main evolutionary lineages of the seed plants. However, there are a few obvious peculiarities. First, the type 2 RIP genes were apparently not transferred into all daughter lineages of the common ancestor of all modern seed plants. Second, type 2 RIP genes are certainly not ubiquitous in all taxa of flowering plants but seem to be patchly distributed. It is certainly true that (many) other homologs remain to be discovered. However, genome sequencing clearly demonstrated that they are absent from, e.g., *Arabidopsis thaliana*,

Table 1 Summary of the documented occurrence of the four different types of ribosome inactivating proteins within the Embryophyta (terrestrial plants)

Anthocerotophyta (hornworts)		No sequences found
Bryophyta (mosses)		No sequences found; A and B domain are absent from the genome of *Physcomitrella patens*
Marchantiophyta (liverworts)		Several proteins comprising two in tandem arrayed ricin B domains are expressed in *Marchantia polymorpha*. No expressed protein with a RIP domain could be retrieved.
Tracheophyta (vascular plants)	Lycopodiophyta (club mosses)	No sequences found; A and B domain are absent from the genome of *Selaginella moellendorffii*

	Euphyllophyta	Moniliformopses
		    Equisetophyta (horsetails): no sequences found
		    Filicophyta (ferns): no sequences found
		
		Spermatophyta (seed plants)
		    Coniferophyta: no sequences found
		    Ginkgophyta: no sequences found
		    Cycadophyta: no sequences found
		    Gnetophyta: type AB from *Gnetum gnemon*
		    Magnoliophyta (flowering plants): A, AB, B, AC and AD*

*See Table 2 for a detailed overview

P. trichocarpa, *Medicago truncatula*, *Glycine max*, *Vitis vinifera*, *Carica papaya*, *Oryza sativa,* and *B. distachyon*. Third, within a given taxon/family, type 2 RIPs might be confined to a single species/genus. For example, apart from two *Abrus* species, type 2 RIPs were not found in any other legume species (despite the fact that the legume family has been extensively explored for the occurrence of lectins). Fourth, in some species, type 2 RIP genes were (strongly) amplified. For example, genome analysis revealed that the genome of *R. communis* contains at least eight type 2 RIP genes. Similarly, the expression of complex mixtures of type 2 RIP in *Sambucus* and *Viscum* species can only be explained by the occurrence of multiple genes.

4.2 Overall Phylogeny of Type 2 RIPs

To further corroborate the evolutionary relationships a phylogenetic analysis was made of all type 2 RIPs for which a (near) complete sequence is available or can be assembled from EST sequences. As shown in Fig. 1 the dendrogram of the type 2 RIPs does not reflect the phylogeny of the species in which they occur but exhibits several obvious anomalies. Only the proteins found in *Sorghum* sp., *S. officinarum*, *Z. mays,* and *P. virgatum* form a distinct clade with a "normal" phylogeny. All other Liliopsida proteins (i.e., those from *I. hollandica*, *E. guineensis,* and *P. multiflorum*) are placed in two different branches together with homologs from unrelated Eudicots/Magnoliids. The most striking anomaly concerns the type

Table 2 Summary of the documented occurrence of the four different types of ribosome inactivating proteins in Magnoliophyta (flowering plants)

Eudicotyledons	Stem Eudicotyledons	Ranunculales			Ranunculaceae	AB*
	Core Eudicotyledons	Asterids	Campanulids		Araliaceae	AB
					Apiaceae	A
					Asteraceae	AB
					Adoxaceae	AB, B
			Ericales		Actinidiaceae	AB, B
					Ebenaceae	B
					Polemoniaceae	AB
					Theaceae	AB, B
			Lamiids		Lamiaceae	A
		Caryophyllales			Aizoaceae	A
					Amaranthaceae	A
					Caryophyllaceae	A
					Nyctaginaceae	A
					Phytolaccaceae	A
		Santalales			Loranthaceae	AB
					Olacaceae	AB
		Rosids	Eurosids I	Cucurbitales	Cucurbitaceae	AB, A
				Fabales	Fabaceae	AB
				Malpighiales	Euphorbiaceae	AB, A, B
					Passifloraceae	AB
					Salicaceae	A, B
				Rosales	Cannabaceae	A
					Rosaceae	A, AB
			Eurosids II	Malvales	Malvaceae	AB
				Sapindales	Sapindaceae	AB
Liliopsida	Asparagales				Agavaceae	A
					Asparagaceae	A
					Iridaceae	AB, A
					Ruscaceae	AB, A
	Commelinids	Arecales			Arecaceae	AB
		Poales			Bromeliaceae	A
			Poaceae	BEP clade	Bambuseae	A
					Oryzeae	A, AD
					Pooideae	A, AB, AC, AD
				PACCAD clade	Panicoideae	A, AB, AC
Magnoliids	Laurales				Lauraceae	AB

*A: type 1; AB: type 2; AC: type 3; AD: new chimeric RIP; B: lectinic B chain
For more details on taxonomic distribution of ribosome-inactivating proteins, please see Supplementary data 2 on http://www.molecularbiotechnology.ugent.be/publications/VanDamme2010A/

2 RIPs from *P. multiflorum* (Van Damme et al. 2000), which are placed in the same cluster as the *Sambucus nigra* proteins. Evidently, the anomalous phylogeny raises some questions with regard to the evolution of type 2 RIPs in flowering plants. To check whether the anomalies might be due to pronounced differences between the evolution of the RIP and the lectin domain, the same phylogenetic analysis was made using sequences of the respective A and B-chains. Though these analyses yielded slightly different results, major anomalies persisted (results not shown). For example, both the A and B-chain of the *P. multiflorum* proteins were invariantly placed in the *Sambucus* cluster.

Summarizing one can conclude that the origin of the type 2 RIP family predates the common ancestor of the Gnetophyta and the Magnoliophyta. The ancestral gene
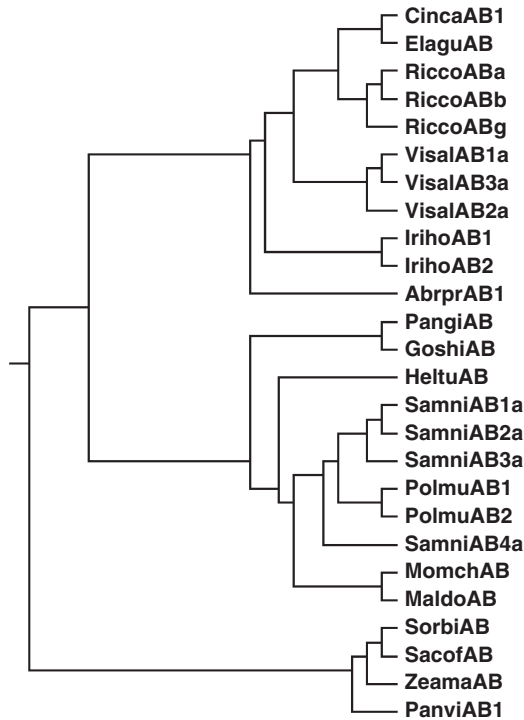
Fig. 1 Phylogenetic analysis of the currently identified type 2 RIPs. Only entries for which complete sequences are available are included. To reduce the complexity of the dendrogram a number of highly similar paralogs from a single species or highly similar orthologs from related species are not included. For more details on the sequences of the ribosome-inactivating proteins and the codes used for each RIP, please see Supplementary data 1 and 3 on http://www.molecularbiotechnology. ugent.be/publications/VanDamme2010A/

was vertically transmitted into the daughter lineages but during further evolution it purged from the genome of most Spermatophyta. This process of gene loss still occurred in evolutionary recent terms and might possibly continue today (as is suggested by the patchy distribution in, e.g., the Poaceae family). The retained genes evolved along with the evolution of the seed plants but it seems that some events took place that eventually resulted in a few obvious but unexplained phylogenetic anomalies.

4.3 Special Evolutionary Events: Gene Amplification and Generation of Type A and Type B Proteins from Genuine Type 2 RIPs

Apart from the general scheme described above, the molecular evolution of type 2 RIPs includes some peculiar events. In several unrelated species gene amplification

has taken place. The occurrence of multiple genes was already inferred from the fact that some plants like, e.g., *Sambucus* sp and *Viscum album* express complex natural mixtures of "isoforms" (Girbés et al. 2004; Mishra et al. 2004). Full details about such a gene family were provided by complete sequencing of the castor bean genome. This genome contains at least eight genuine type 2 RIP genes and several (at least eight) pseudogenes (with an incomplete or interrupted ORF). In addition, a type 2 RIP gene could be retrieved that encodes a protein with a heavily truncated B-chain (only the first 44 amino acid residues are left). Interestingly, a perfectly matching EST sequence is deposited, which indicates that the truncated protein might be expressed. Though there is no experimental evidence yet for the presence of the corresponding protein in castor bean tissue, the truncated AB gene illustrates that a type 1 RIP can be generated from a genuine type 2 RIP through the deletion of the B domain. This observation is not merely anecdotal but has important consequences for what concerns the molecular evolution of type 1 RIPs. As was already suggested on several occasions (Peumans et al. 2001; Van Damme et al. 2001) some type 1 RIPs (e.g., from *I. hollandica*) (Van Damme et al. 1997a) are more closely related, indeed, to type 2 RIPs from the same or a related species than to any other (genuine) type 1 RIP. Analysis of the castor bean genome now provides a firm basis for the formerly predicted "domain-deletion" origin of some type 1 RIPs and as such urges the reassessment of the evolutionary origin of the whole group of type A proteins (which is discussed in detail in a separate section).

Comprehensive biochemical and molecular analyses demonstrated that *Sambucus* sp. also express a very complex set of both genuine and truncated type 2 RIP genes and revealed that within this taxonomic group an evolutionary event took place whereby a genuine type 2 RIP (called SNA-V) was converted in a type B protein (SNA-II) through a deletion of the RIP domain (Van Damme et al. 1996, 1997b). To check whether the origin of the other identified type B proteins relies on a similar mechanism, a phylogenetic analysis was made of the sequences of these proteins and the corresponding sequences of the B-chain of type 2 RIPs. According to the dendrogram shown in Fig. 2, the B-type proteins found in *P. trichocarpa* cluster with the type 2 RIP from *Adenia volkensii* (which belongs to the same order Malpighiales as poplar) indicated that they are derived, indeed, from a genuine type 2 RIP through deletion of the A domain – an event that most probably took place after the Passifloraceae and Adoxaceae diverged from the Euphorbiaceae. Since the parent type 2 RIP gene is no longer present in poplar, it must have been purged from the genome during the evolution of the lineage Malpighiales that led to modern poplars. At the same time, the original gene encoding a B-type protein was amplified by two in tandem duplications followed by a region/chromosome or genome wide duplication. In addition to B-type genes, the poplar genome contains a gene encoding a protein consisting of a single ricin-B domain (or a half B-chain). The origin of this gene is not clear. It exhibits the highest similarity with the C-terminal domain of the B-type proteins in poplar but the sequence identity is relatively low (approximately 35% within the ricin-B domain) indicating that there is most likely no direct evolutionary link. Possibly the single ricin-B domain gene
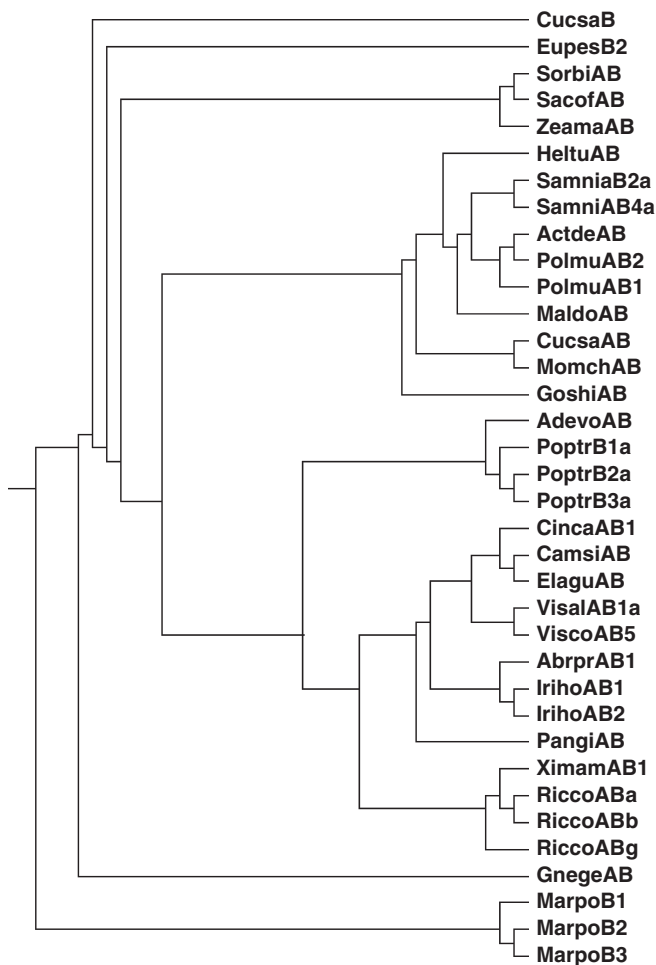
Fig. 2 Phylogenetic analysis of all identified plant B-chain proteins and a selected set of the B-chains of type 2 RIPs. To reduce the complexity of the dendrogram a number of highly similar type 2 RIP paralogs from a single species or highly similar orthologs from related species are not included

results from a different evolutionary event whereby both the A domain and the first half of the B domain were deleted from a genuine type 2 RIP gene.

Taking into consideration the obvious evolutionary origin of the *Sambucus* and *Populus* B-chain proteins, the position in the dendrogram of the homologs from *C. sativus* and *E. esula* also can be reconciled with a similar domain-deletion event. However, in these two cases the respective evolutionary events took place most likely in a distant past.

A final important conclusion to be drawn from the dendrogram shown in Fig. 2 concerns the cytoplasmic B-chain proteins expressed in *M. polymorpha*. These proteins are apparently unrelated to any other ricin-B protein found in plants and

accordingly are not likely candidate ancestors of the ricin-B domain found in type 2 RIPs and related proteins. As matter of fact, the liverwort proteins share little sequence similarity with any known protein (the best match is an α-galactosidase from the Actinomycete bacterium *Catenulispora acidiphila* DSM (gb|EEN35459.1|) and hence might represent a separate subgroup of the ricin-B family).

4.4 What is the Origin of Type 2 RIP Genes?

Hitherto, no homologs of the type 2 RIPs have been identified outside the seed plants, which strongly indicates that this chimeric protein was developed within the lineage leading to the modern Spermatophyta (or alternatively was acquired by non vertical inheritance from an unidentified organism that lost the corresponding gene or became extinct). All that can be stated with certainty is that the ancestor of today's type 2 RIPs already existed before the Gnetophyta and Magnoliophyta lineages diverged from each other. Evidently, the chimeric structure must result from an evolutionary event whereby a RIP domain was fused to a duplicated ricin-B domain.

4.4.1 Origin of the B-Chain

On the basis of the widespread distribution over all major prokaryotic and eukaryotic taxa, one can reasonably assume that the ricin-B fold was developed by bacteria very early in evolution and transmitted into all major eukaryotic lineages. Unfortunately, because of the apparent absence of the (sequenced) genomes of Chlorophyta, it is impossible to trace the origin of the plant ricin-B domain back to the common ancestor of plants and other eukaryotes. Sequence comparisons revealed that the B-chain of the type 2 RIPs shares the highest similarity with the (double ricin-B domain) carbohydrate-binding part of an (extracellular or secreted) β-glycosidase-like glycosyl hydrolase (gb|EEN27866.1|) and an α-L-arabinofuranosidase B family protein (gb|EEN23780.1|) from the Actinomycete bacterium *C. acidiphila*. Moreover, all the cysteine (Cys) residues that stabilize (by four disulfide bonds) the rigid fold of the B-chain of the plant type 2 RIPs are also present in these bacterial sequences whereas virtually all other eukaryotic ricin-B domains lack disulfide bonds. The latter fact is not surprising because apart from type 2 RIPs and related proteins most eukaryotic proteins with ricin-B domains are synthesized on free ribosomes in the reducing cytoplasm where disulfide bonds are not usually formed. Hence, all evidence suggests that at a given time (at least predating Gnetophyta and Magnoliophyta lineages) in the evolution of the lineage Spermatophyta, a plant acquired a gene encoding a protein consisting of a duplicated Cys-rich ricin-B domain by lateral transfer from a bacterium. Possibly, the cytoplasmic ricin-B proteins found in *M. polymorpha* were also acquired by a similar lateral transfer from a bacterium, but in this case the bacterial gene encoded a cytoplasmic (Cys-poor) protein.

4.4.2   Origin of the A-Chain

The A domain is far less common than the ricin-B domain. Apart from plants, it seems to be confined to bacteria and viruses of the (entero) bacteriophage group. Moreover, within the bacteria the prokaryotic homolog of the type 1 RIP (called Shiga and Shiga-like toxins A component) is rather rare as it is found only in a small number of species. Proteins were identified in not more than 13 species, including *Streptomyces coelicolor*, 11 Enterobacteriaceae sp., and *Acinetobacter haemolyticus*. This very narrow taxonomic distribution (especially when compared to the widespread occurrence in plants) raises some questions with regard to the presumed bacterial origin. An alternative explanation might be that the RIP domain was developed by plants and acquired by some bacteria through lateral gene transfer. The latter (admittedly speculative) hypothesis is supported by the fact that the target of the Shiga toxins (globotriaosylceramide, a typical animal glycolipid), was developed later in evolution than the RIP domain, which implies that there was no selective pressure for bacteria to develop the RIP domain before it was already present in plants. Irrespective of the true origin, the present taxonomic distribution indicates that the A domain was developed or acquired by an early seed plant before the Gnetophyta and Magnoliophyta lineages diverged from each other. No type 1 RIP could be found in the transcriptome of *G. gnemon* but the limited number of sequences does not allow the conclusion that the corresponding gene is absent. Unfortunately, this implies that no direct link can be made between the A chain of the *G. gnemon* type 2 RIPs and a possible type 1 RIP from the same species.

5   Molecular Evolution of Type 1 RIPs

Type 1 RIPs are not only more common in seed plants they also exhibit a much higher (sequence) heterogeneity than type 2 RIPs. In the past, type 1 RIPs were usually subdivided in three groups (Van Damme et al. 2001). The first group comprises the "classical" type 1 RIPs found in numerous dicotyledons (e.g., in Cucurbitaceae, Phytolaccaceae, and Amaranthaceae species). All these RIPs are synthesized on the ER and follow the secretory pathway to their final subcellular destination (vacuole/extracellular space). The second group is exemplified by the RIPs that were isolated from wheat, barley, and some other grasses. These Poaceae RIPs are synthesized without signal peptide and presumably reside in the cytoplasm. Members of the third group are also found in Poaceae and more precisely in *Z. mays* and related (Panicoideae) species (Walsh et al. 1991; Hey et al. 1995). What is special about this group is the fact that they are synthesized (on free ribosomes) as inactive precursors that are converted into an enzymatically active form through the proteolytic removal of a short peptide at both the N- and C-terminal end, and the excision of an internal peptide. As a result the active RIP consists of two different polypeptides and accordingly is also referred to as a "two-chain" type 1 RIP.

Most attempts to elucidate the phylogenetic relationships were based on the assumption that the family of type 1 RIPs is, notwithstanding the obvious heterogeneity, monophyletic. However, as earlier work with type 1 and type 2 RIPs from *I. hollandica* (Van Damme et al. 1997a) and phylogenetic analyses of type 1 and type 2 RIPs already indicated (Van Damme et al. 2001), sequencing of the castor bean genome provided firm evidence that type 1 RIPs are generated from parent type 2 RIP genes. Since these two examples might be indicative of a more general evolutionary mechanism the phylogeny of type 1 RIPs was reassessed using a novel approach. At the same time the phylogeny was refined by incorporating recent sequence data. Taking into account that the type 1 RIPs from Poaceae differ in several aspects from those found in dicots and monocots other than Poaceae, the two groups were first analyzed individually. After establishing the phylogenetic relationships within each group, a comprehensive analysis was made of all plant RIPs using a selected set of sequences of both groups. Finally, to corroborate the possible link with bacterial RIPs, a selection of bacterial sequences was included in the final analysis.

5.1   Dicots and Monocots Other Than Poaceae

Due to the large number of sequences, it was virtually impossible to make a phylogenetic analysis in a single step. Therefore, a reiterative process was followed whereby smaller sets of sequences from taxonomically related species were analyzed. To reduce the complexity, highly similar paralogs from single species and/or orthologs from related species were omitted and the limited set of sequences combined with similarly reduced sets from other taxonomic groups. Using this approach, a fairly accurate dendrogram could be generated of the sequences of all type 1 RIPs except those from the Poaceae species. As shown in Fig. 3, the resulting dendrogram comprises three distinct clades. The first clade, which groups all Caryophyllales type 1 RIPs comprises two side branches corresponding to two distinct forms. Since both forms are found in *Beta vulgaris* (Amaranthaceae) and *Mesembryanthemum crystallinum* (Aizoaceae), they most probably result from a gene or genome duplication that predates the separation of the different Caryophyllales families. The second major clade comprises all Eurosids I sequences and consists of two-side branches with a cluster of Euphorbiaceae and Cucurbitaceae proteins, respectively. The only documented type 1 RIPs from Rosaceae (from *Prunus* sp.) and Ericales (from *C. sinensis*) are also placed in the Cucurbitaceae cluster. The third clade is well separated from the two other groups – the type 1 RIPs from the dicot *P. trichocarpa* and the monocots *Muscaria*, *Asparagus,* and *Ophiopogon* – and thus covers a broader taxonomic range than the other clades. The main conclusion that can be drawn is that apart from a few exceptions the dendrogram of the type 1 RIPs reflects the taxonomy of the dicots. However, the dendrogram also reveals the occurrence of a second more distant group found in both dicots and monocots.
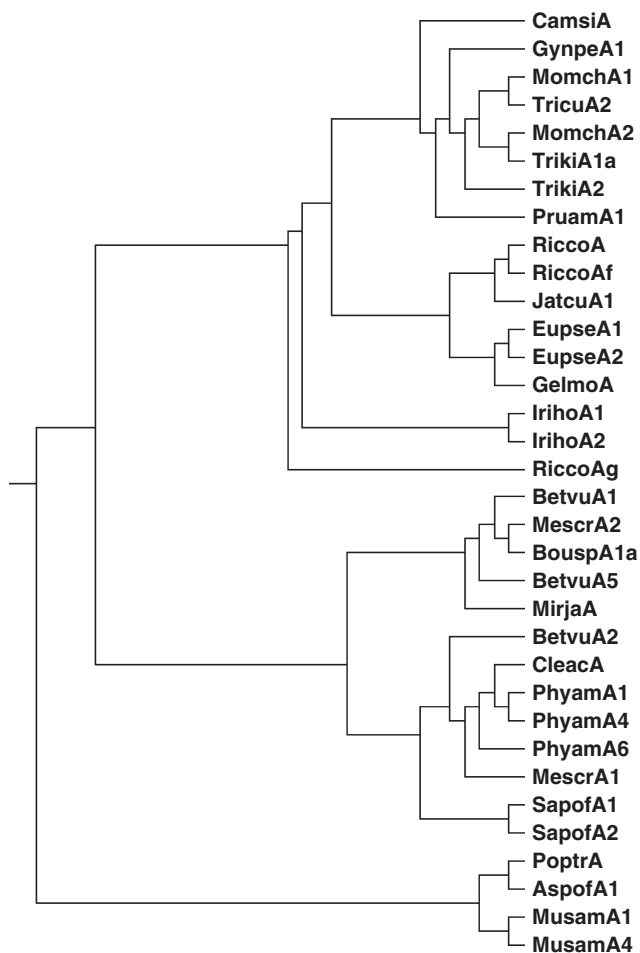
Fig. 3 Phylogenetic analysis of type 1 RIPs from dicots and monocots other than Poaceae. Highly similar paralogs from a single species or highly similar orthologs from related species are not included

To corroborate which groups of type 1 RIPs are the products of domain-deletion (of AB genes), the phylogenetic analysis was extended by incorporating the A domain (including signal peptide) of type 2 RIPs (Fig. 4). The resulting dendrogram leaves no doubt that the type 1 RIPs from iris originated from a conspecific type 2 RIP in an evolutionary recent past. The same applies to the type 1 RIPs from Rosids (*Malus* and *Momordica*), but in this case the domain loss apparently predates the divergence of the Cucurbitales and Rosales. Reasoning along the same line, it seems likely that the Euphorbiaceae type 1 RIPs also evolved from type 2 RIPs in a more distant past. The dendrogram leaves some
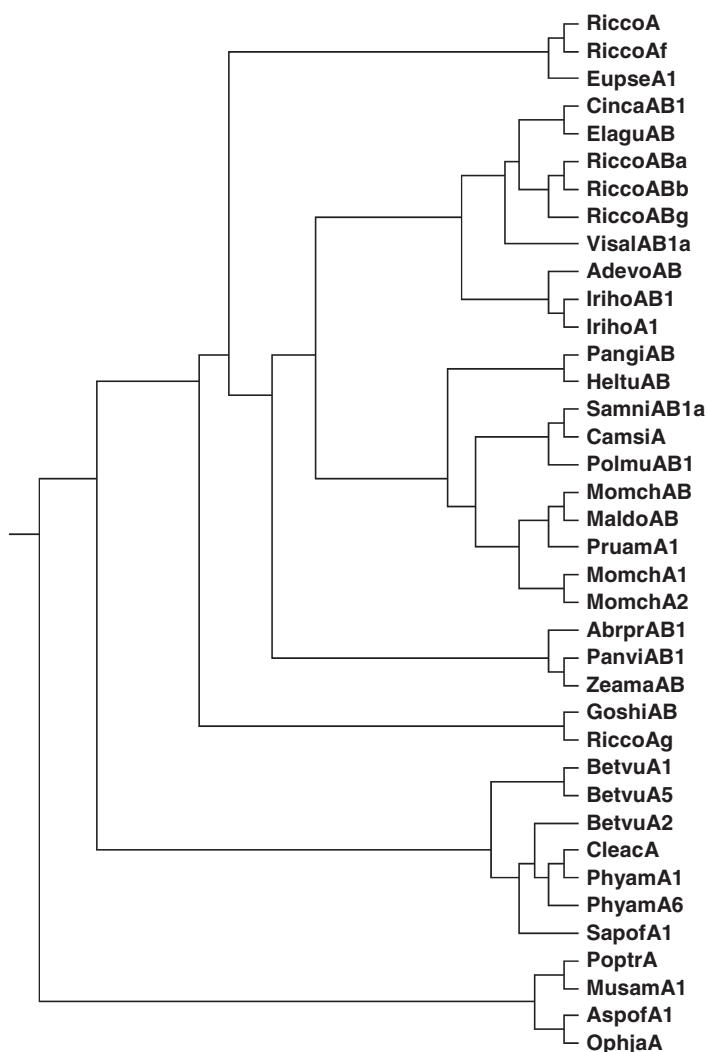
Fig. 4 Preliminary phylogenetic analysis of a selected set of sequences of type 1 RIPs from dicots and monocots other than Poaceae and the N-terminal domain (including signal peptide) of type 2 RIPs. Only sequences relevant for the issue of the B domain loss origin of most dicot type 1 RIPs are included

uncertainty with regard to the origin of the homologs from Caryophyllales and *Clerodendron* (but as discussed below, they were also derived in a more distant past from a type 2 RIP gene). This implies that only the sequences in the "aberrant" clade with *Populus*, *Muscaria,* and *Ophiopogon* are derived from a genuine type 1 RIP.

5.2   Poaceae Type 1 RIPs

It has been known for more than a decade that many Poaceae species express
complex mixtures of type 1 RIPs. However, recent genome and transcriptome
sequencing data revealed that the complexity inferred from biochemical and molec-
ular analyses is still an underestimation of the total RIP gene complement. The
following examples illustrate the complexity of the RIP gene family in Poaceae.

5.2.1   O. sativa

In a recent study, Jiang et al. (2008) reported that the "RIP domain family" in
O. sativa comprises 31 different members. A careful reanalysis of the genome
indicated that six presumed RIP genes are pseudogenes or part of a transposon.
Moreover, at least three additional genes could be identified (which are not anno-
tated yet) yielding a total number of 28 genes. Phylogenetic analyses indicated that
the rice genes cluster in four major clades (Fig. 5) and confirmed the conclusion
drawn by Jiang et al. (2008) that the expansion of the RIP gene family is primarily
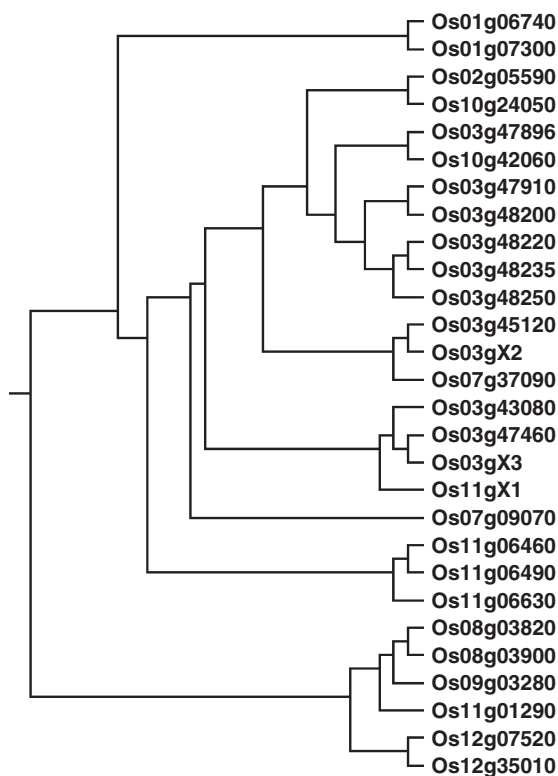based on genome-wide duplications and to a lesser extent on tandem duplications.

Fig. 5 Phylogenetic analysis
of the *Oryza sativa* RIP
family. Several annotated
sequences have been
corrected. For Os11g06460
and Os11g06490 (which are
AD type chimers) sequences
corresponding to the A
domain were used

5.2.2 Andropogoneae: *Z. mays* and *Sorghum bicolor*

Similar analyses indicated that the RIP gene family in the genomes of *Z. mays* and *S. bicolor* is less extended than in *O. sativa,* but is still fairly complex. However, unlike rice these two Andropogoneae species possess a genuine type 2 RIP gene (Figs. 6–9). In addition, a set of two type AC chimeric genes occurs in the genome of maize but not in that of sorghum. The dendrogram of the combined maize and sorghum gene families consists of two major clades one of which comprises three side branches. Apart from the maize AC type gene, all other types of genes are present in both species but there are obvious differences in the number of genes in
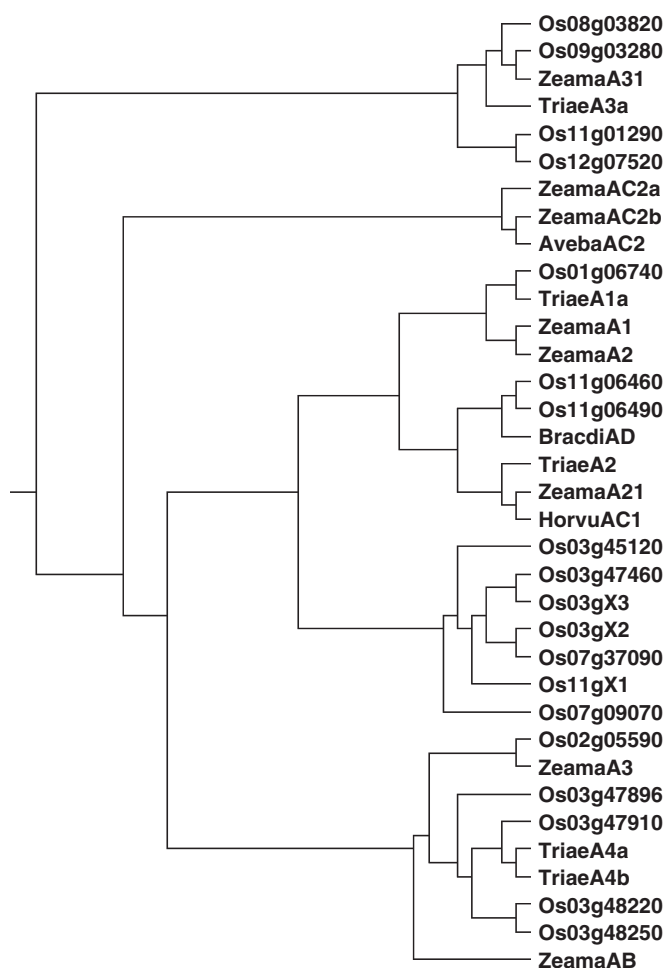

Fig. 6 Phylogenetic analysis of the RIP gene family within the family Poaceae. To reduce the complexity of the dendrogram three limited sets of sequences from *Triticum aestivum*, *Zea mays,* and *Oryza sativa* were combined

each branch. This indicates that the expansion of the RIP gene family in maize and sorghum was differentially affected by gene duplication.

5.2.3 Pooideae

Hitherto *B. distachyon* is the only Pooideae species for which a complete family of RIP genes can be retrieved.[2] The RIP gene family consists of four different type 1 and one type AD RIP gene and hence is far less complex than in rice, maize, and sorghum.

Though no genomic data are available for any other Pooideae species the vast amount of transcriptome data leave no doubt that, e.g., wheat and barley possess extended RIP families. The same applies most probably to *Secale cereale, Leymus cinereus, Aegilops speltoides,* and *Pseudoroegneria spicata.*

To corroborate the evolution of the RIP gene families within the Poaceae family the overall phylogeny was analyzed using three sets of sequences from rice, maize, and wheat, respectively. To reduce the complexity highly similar orthologs/paralogs (clustering in a single branch or side branch) were omitted (except one or two). As shown in Fig. 6, the dendrogram of the combined Poaceae sequences closely resembles that of the rice RIP gene family except that a branch with a type 2 RIP is introduced. This confirms the conclusion by Jiang et al. (2008) that the RIP gene family has largely evolved in parallel to species evolution within the family Poaceae. However, there seem to be important differences in what concerns the contribution of tandem duplications in different representatives. Moreover some species possess genes that are absent in others. For example, the type AB and AC genes present in and expressed by the maize genome are definitely absent from the rice genome. This finding indicates that the (still ongoing) evolution of the RIP gene family within the family Poaceae is to a certain extent determined by gene loss.

5.2.4 Relationships between the RIPs from Poaceae and Other Seed Plants

Though the Poaceae type 1 RIPs are usually regarded as a separate group, one can reasonably assume that they are somehow related to homologs from both other monocots and dicots. Therefore, a phylogenetic analysis was made of the combined sequences of type 1 RIPs from both dicots and all monocots (and supplemented with the A domain of type 2 RIPs). According to the results shown in Fig. 7, the issue of overall evolution of type 1 and type 2 RIP genes needs a thorough update. All sequences, apart from an orphan sequence found in wheat, cluster in two clades. One of these clades comprises all known dicot type 1 RIPs except that from poplar and all type 2 RIP sequences, whereas the second clade groups all monocot type 1

---

[2]These sequence data were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/.
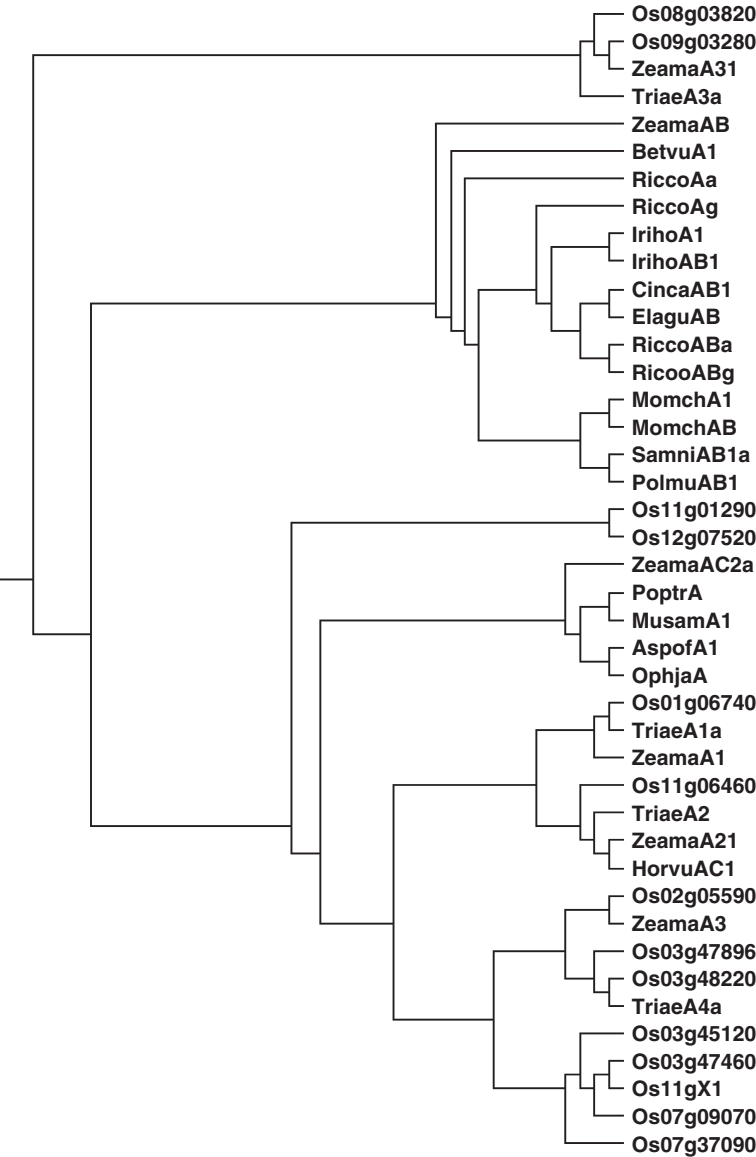
Fig. 7 Phylogenetic analysis of all plant RIPs

RIPs and the poplar sequence. This overall pattern clearly indicates a double phylogenetic origin. The second clade most probably represents a lineage that starts with an ancestral type A gene and leads to the monocot and poplar genes (and hence can be considered "primary type 1 RIPs" whereas the first clade represents a line of "secondary" type 1 RIPs that are derived from the type 2 RIP lineage through multiple B domain loss events.

6 What is the Relationship between Plant and Bacterial RIPs?

Several bacterial species (and bacteriophages) possess genes encoding one or more proteins with an RIP domain. Though it is generally accepted that the bacterial and plant sequences have a common origin, the exact relationships remain to be elucidated. Therefore, it seemed worthwhile to include the bacterial proteins in the same dendrogram as the plant proteins (Fig. 8). Only a limited number of bacterial proteins with an RIP domain have been identified. The best known are the Shiga toxin A subunit and (closely) related proteins found in *Escherichia coli*, *Shigella* sp., and some other Enterobacteria. All these proteins can be considered homologs of plant type 1 RIPs. In addition, a larger chimeric gene with an N-terminal RIP domain linked to an unrelated domain could be identified in *Micromonospora* sp. When added to the set of plant sequences the bacterial proteins do not cluster in a single prokaryotic clade but are placed at different positions. For example, the A domain of the *Micromonospora* protein forms a small side branch together with one of the type 1 RIP from wheat whereas the other bacterial proteins represent two small side branches of the type 2 RIP clade. Additional evidence against a monophyletic origin of the bacterial sequences comes from the observation that the *Micromonospora* protein is not retrieved by BLAST searches using, e.g., the Shiga toxin as a query, but is readily detected with the *Muscaria* type 1 RIP sequence. This, taken together with the narrow and patchly taxonomic distribution, makes it difficult to reconcile with a prokaryotic origin and therefore strengthens the idea that a few bacteria acquired the RIP domain by lateral transfer from a plant.

7 Chimeric RIPs Other Than Type 2 RIPs

Besides type 1 and type 2 RIPs two additional types of chimeric RIP proteins have been identified in Poaceae species.

7.1 JIP60 and Other Type AC Chimeric RIPs

A chimeric RIP protein was isolated from jasmonate treated barley leaves (Chaudhry et al. 1994). The so-called 60 kDa jasmonate-induced protein (JIP60) consists of an N-terminal A domain fused to an unrelated domain with no known function (which will further be referred to as the C domain) and was classified as a type 3 RIP (or AC-type). Hitherto, no homologs have been isolated but genome and transcriptome data revealed the occurrence of similar proteins in *Z. mays* and *Avena barbata*. Since there is no evidence for the presence outside the family Poaceae, the type AC RIP originated most likely within this family. To trace the origin of the RIP domain of these AC proteins the sequences of their respective A domains were
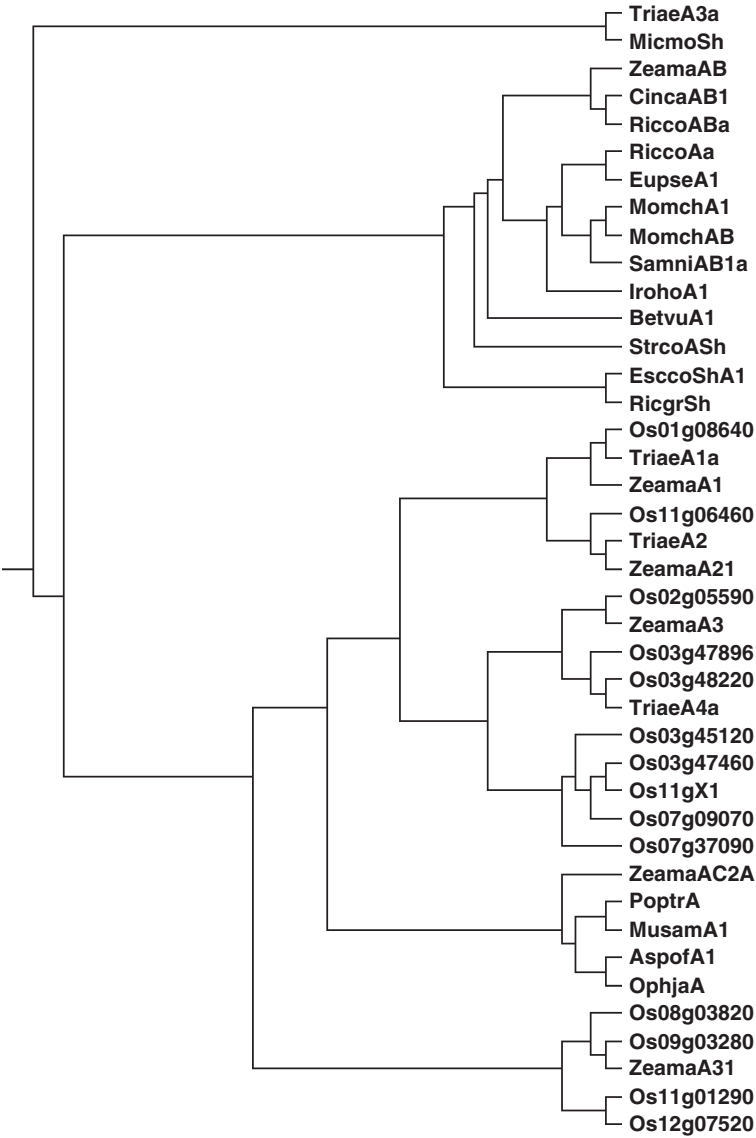
Fig. 8 Phylogenetic analysis of combined plant and bacterial RIP sequences

incorporated in a representative set of Poaceae type 1 RIPs for a phylogenetic analysis. As shown in Fig. 9 the RIP domains of AC proteins do not form a separate cluster but the A domain of HorvuAC (JIP60) and these from *Z. mays*/*A. barbata* are divided over two major branches. Moreover, one of the maize proteins is apparently more closely related to the oat homolog than to its conspecific paralog.

Fig. 9 Phylogenetic analysis of the C domain of the type AC RIPs, the D domain of type AD RIPs, and C and D proteins
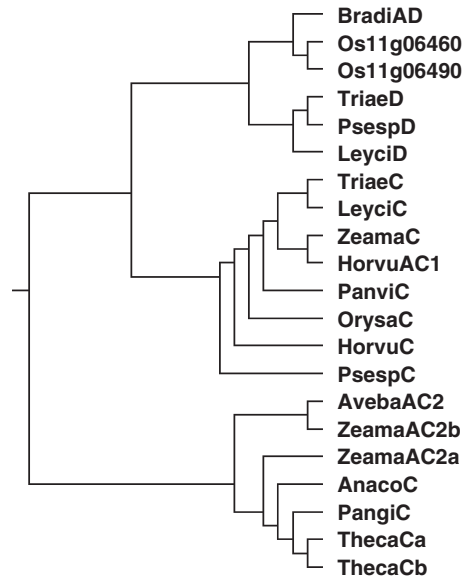
The overall shape of the dendrogram and the obvious phylogenetic anomaly (one would expect that oat and barley cluster together) are indicative of two different gene fusion events: one that led to the AC1 type (as in barley) and another that yielded the oat and maize AC2 type. To check the presumed double origin a parallel phylogenetic analysis was made of the C domains found in the AC chimers and their parent genes. BLAST searches revealed that most but certainly not all Poaceae species possess (expressed) genes encoding a protein equivalent to the C domain of the type AC RIPs. For example, the C protein is found in wheat, barley, rice, maize, and *Panicum* but is absent in the *Sorghum* genome. Homologs are also expressed in other monocots (e.g., pineapple) and in some dicots (e.g., *P. ginseng* and *Theobroma cacao*). A phylogenetic analysis of all these sequences confirms that the barley type AC1 protein on the one hand and the maize and oat homologs on the other hand result most probably from two independent domain fusion events. The high sequence similarity between the barley AC1 protein and the C proteins found in other Poaceae indicates that the JIP60 chimer arose in an evolutionary recent past through a fusion event between a type 1 RIP gene of the TriaeA2 clade and a C domain gene. Since there is no information about the presence or absence of genuine JIP60 orthologs in other Poaceae, it is difficult to date the time point of the fusion event, but it might have taken place (relatively) recently in an ancestral Triticeae species. According to the dendrogram of the C domains, the origin of the AC2 type goes much further back in the time. Moreover, the position of the AC2 proteins in the dendrogram shown in Fig. 9 suggests that a different type 1 RIP gene was involved as in the fusion leading to the AC1 type chimer.

7.2   Chimeric RIP with a C-terminal D Domain

BLAST searches in the *B. distachyon* database (http://www.brachybase.org) yielded a (genomic) sequence sharing a high sequence similarity with the rice proteins Os11g06460 and Os11g06490. A perfectly matching (but short) EST sequence was retrieved indicating that the protein is expressed in *Brachypodium*. A closer examination revealed that the *Brachypodium* protein contains, besides an N-terminal RIP domain, a long terminal extension that shares a reasonably high sequence similarity with an unidentified protein (further referred to as the D protein) expressed in wheat, *P. spicata,* and *L. cinereus*. Using the *Leymus* D sequence as a model the *B. distachyon* genomic sequence could be correctly spliced. Once the exon–intron structure of the *Brachypodum* AD protein was determined the genomic rice sequences encoding Os11g06460 and Os11g06490 were reanalyzed and spliced correctly. Both rice proteins share a high sequence similarity with the *Brachypodium* AD chimer and comprise a C-terminal domain equivalent to the D protein from *Leymus*. The latter protein is still unidentified but shares sequence identity with the C proteins. For example, the C and D proteins share 24% identity concentrated in several well conserved regions. Therefore, the sequences of the AD chimers and the D proteins were included in the same phylogenetic analysis as the AC chimers and sole C proteins (Fig. 9). According to the dendrogram shown in Fig. 9 the A domain involved in the gene fusion leading to the AD chimers belonged to the same clade as the one recruited for the formation of the barley JIP60 AC1-type RIP. Given the occurrence of genuine orthologs in *Oryza* and *Brachypodium* the origin of the AD type RIP predates the division of the BEP clade into the Eahrhartoideae and Pooideae lineages.

8   Conclusions

The availability of novel genome and transcriptome data allowed updating the molecular evolution of the RIP family in plants (Fig. 10). All evidence suggests that the RIP domain itself was developed in plants before the Gnetophyta and Magnoliophyta lineages diverged from a common ancestor. Then the ancestral RIP domain followed two separate routes. A first route led directly to a subset of modern type 1 RIPs (primary type 1 RIPs). The second route started with a fusion (also before the Gnetophyta and Magnoliophyta lineages separated) of the RIP domain to a sugar-binding domain, which might be acquired by lateral transfer from a bacterium, and subsequently gave rise to (1) modern type 2 RIPs, (2) multiple lines of (secondary) type 1 RIPs through several domain B deletion events and (3) several lines of type B lectins through domain A deletion events. In many species, the RIP genes were purged from the genome whereas other species developed a whole set of RIP genes by gene and genome wide amplifications. Within the family Poaceae the original type 1 RIPs gave rise to (1) a complex set of type 1 RIPs
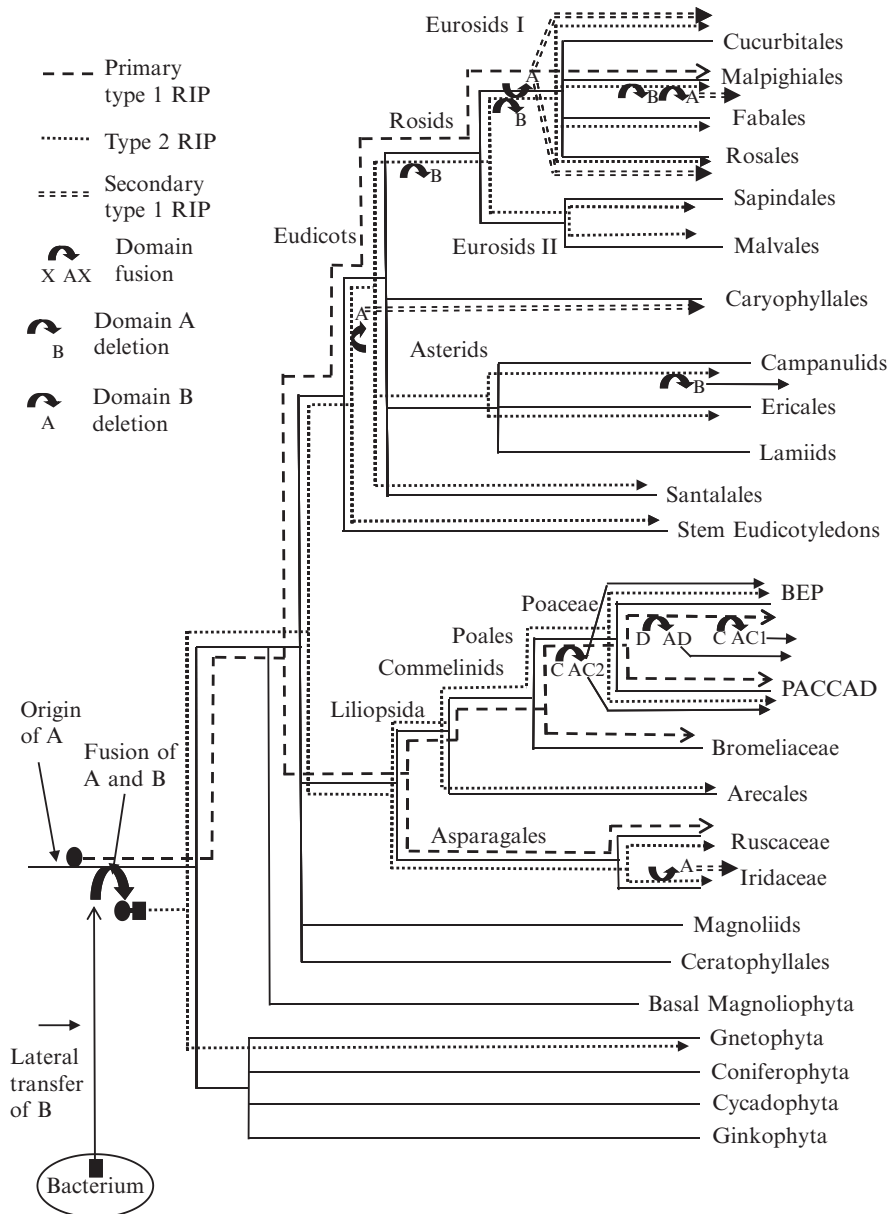
Fig. 10 Overall scheme of evolution of the RIP family in plants based on documented occurrence and phylogeny of type 1, type 2, and type AC and AD RIPs

through multiple gene and genome amplifications and (2) at least two different chimeric forms by two independent fusion events of a type 1 RIP to a C- and D-type protein, respectively. All evidence suggests that in the family Poaceae the RIP gene

family is still evolving by different mechanisms. A final remark concerns the origin of the bacterial RIP domain. Bacteria most probably did not develop their own RIP domain but acquired it through (multiple) lateral gene transfers from a plant.

References

Barbieri L, Battelli MG, Stirpe F (1993) Ribosome-inactivating proteins from plants. Biochim Biophys Acta 1154:237–282

Chaudhry B, Müller-Uri F, Cameron-Mills V, Gough S, Simpson D, Skriver K, Mundy J (1994) The barley 60 kDa jasmonate-induced protein (JIP60) is a novel ribosome-inactivating protein. Plant J 6:815–824

Girbés T, Ferreras JM, Arias FJ, Stirpe F (2004) Description, distribution, activity and phylogenetic relationship of ribosome-inactivating proteins in plants, fungi and bacteria. Mini Rev Med Chem 4:461–476

Hey TD, Hartley M, Walsh TA (1995) Maize ribosome-inactivating protein (b-32). Homologs in related species, effects on maize ribosomes, and modulation of activity by pro-peptide deletions. Plant Physiol 107:1323–1332

Jiang SY, Ramamoorthy R, Bhalla R, Luan HF, Venkatesh PN, Cai M, Ramachandran S (2008) Genome-wide survey of the RIP domain family in *Oryza sativa* and their expression profiles under various abiotic and biotic stresses. Plant Mol Biol 67:603–614

Masuda S, Sakuta C, Satoh S (1999) cDNA cloning of a novel lectin-like xylem sap protein and its root-specific expression in cucumber. Plant Cell Physiol 40:1177–1181

Mishra V, Sharma RS, Yadav S, Babu CR, Singh TP (2004) Purification and characterization of four isoforms of Himalayan mistletoe ribosome-inactivating protein from *Viscum album* having unique sugar affinity. Arch Biochem Biophys 423:288–301

Palmer JD, Soltis DE, Chase MW (2004) The plant tree of life: an overview and some points of view. Am J Bot 91:1437–1445

Peumans WJ, Hao Q, Van Damme EJM (2001) Ribosome-inactivating proteins from plants: more than RNA *N*-glycosidases? FASEB J 15:1493–1506

Stirpe F, Battelli MG (2006) Ribosome-inactivating proteins: progress and problems. Cell Mol Life Sci 63:1850–1866

Stirpe F, Licastro F, Morini MC, Parente A, Savino G, Abbondanza A, Bolognesi A, Falasca AI, Rossi CA (1993) Purification and partial characterization of a mitogenic lectin from the latex of *Euphorbia marginata*. Biochim Biophys Acta 1158:33–39

Van Damme EJM, Barre A, Rougé P, Van Leuven F, Peumans WJ (1996) Characterization and molecular cloning of SNAV (nigrin b), a GalNAc-specific type 2 ribosome-inactivating protein from the bark of elderberry (*Sambucus nigra*). Eur J Biochem 237:505–513

Van Damme EJM, Barre A, Barbieri L, Valbonesi P, Rougé P, Van Leuven F, Stirpe E, Peumans WJ (1997a) Type 1 ribosome-inactivating proteins from Iris (*Iris hollandica* var. Professor Blaauw) bulbs: characterization and molecular cloning. Biochem J 324:963–970

Van Damme EJM, Roy S, Barre A, Rougé P, Van Leuven F, Peumans WJ (1997b) The major elderberry (*Sambucus nigra*) fruit protein is a lectin derived from a truncated type 2 ribosome-inactivating protein. Plant J 12:1251–1260

Van Damme EJM, Peumans WJ, Barre A, Rougé P (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. Crit Rev Plant Sci 17:575–692

Van Damme EJM, Hao Q, Charels D, Barre A, Rougé P, Van Leuven F, Peumans WJ (2000) Characterization and molecular cloning of two different type 2 ribosome-inactivating proteins from the monocotyledonous plant *Polygonatum multiflorum*. Eur J Biochem 267:2746–2759

Van Damme EJM, Hao Q, Chen Y, Barre A, Vandenbussche F, Desmyter S, Rougé P, Peumans WJ (2001) Ribosome-inactivating proteins: a family of plant proteins that do more than inactivate ribosomes. Crit Rev Plant Sci 20:395–465

Van Damme EJM, Lannoo N, Peumans WJ (2008) Plant lectins. Adv Bot Res 48:107–209

Walsh TA, Morgan AE, Hey TD (1991) Characterization and molecular cloning of a proenzyme form of a ribosome-inactivating protein from maize. Novel mechanism of proenzyme activation by proteolytic removal of a 2.8-kilodalton internal peptide segment. J Biol Chem 266:23422–23427