

Computational internal sequence repeats analysis of accelerated evolution and the role of extensins under abiotic and biotic stresses

Ramalingam Jothi^{1,2}
Subbiah Parthasarathy¹
Kulithalai Viswanathan
Krishnamurthy¹

¹Department of Bioinformatics,
School of Life Sciences, Bharathidasan
University, Tiruchirappalli;

²Department of Zoology,
Dharmapuram Gnanambigai
Government Arts College (Women),
Mayiladuthurai, Tamil Nadu, India

Abstract: Extensins are hydroxyproline-rich glycoproteins present in the plant cell wall. Accelerated evolution and the role of extensins involved in abiotic stresses, like gravitational stress and tension wood formation, and biotic stress, like pathogenic resistance, were investigated through computational internal sequence repeats analysis of their sequences. Multiple sequence alignment analysis of extensin-1 and extensin-2, present in both herbs and trees, was used to investigate their role in gravitational stress and tension wood formation. The role of extensins in pathogenic resistance was investigated by showing the existence of circular permutation in both extensin-1 and extensin-2 sequences between plants of *Fabaceae*. It was analyzed through Dot plots, and the study predicted that many partial circular permutations exist between sequences of different length. The clustering study of the internal sequence repeats of extensin family through phylogenetic analysis and reconstruction of patterns of repeat identities between sequences showed that unequal crossover exists in extensin-2. From the existence of unequal crossover nature in the circularly permuted extensin-2, accelerated evolution is indicated by the pattern of repeats in herbaceous plants of the *Fabaceae* family and trees of different taxa, in addition to pathogenic resistance.

Keywords: extensin, abiotic and biotic stresses, tension wood, circular permutation, unequal crossover, pathogenic resistance

Introduction

One of the most notable evolutionary changes in land plants is the development of different kinds of cell walls. During growth, development, environmental stresses and infection, the cell wall is continuously modified by enzyme action.¹ Thus, it is important to study how cell walls are made more flexible locally to result in the differential growth process observed in plant morphogenesis. Proteins localized in cell walls are ubiquitous and usually rich in certain amino acids containing highly repetitive sequence domains. Among these are the hydroxyproline-rich glycoproteins or extensins, arabinogalactan proteins, glycine-rich proteins, and proline-rich proteins.

Extensins are a category of structural, hydroxyproline-rich, highly insoluble glycoproteins found in the cell wall of plants.^{1,2} In general, extensin genes encode proteins with a signal peptide followed by a repetitive region rich in proline (Pro)/hydroxyproline, with the main repeat motif being serine (Pro)₄.^{1,3} Moreover, the serine (Pro)₄ repeat units are typically thought to be characteristics of extensins.³ Generally extensins are distributed in dicots and are less common in monocots. In dicots, extensins are particularly abundant in plants of the *Fabaceae* family, such as *Medicago truncatula*, *Phaseolus vulgaris*, *Cicer arietinum*, *Glycine max*, *Medicago sativa*,

Correspondence: Subbiah Parthasarathy
Department of Bioinformatics, School
of Life Sciences, Bharathidasan University,
Tiruchirappalli 620 024, Tamil Nadu, India
Tel +91 944 353 3095
Fax +91 431 240 7045
Email bdupartha@gmail.com

Pisum sativum, *Trifolium repens*, *Lupinus luteus*, *Vicia faba*, etc. Extensins are also present in other dicots of different taxa, like *Arabidopsis thaliana*, *Gossypium hirsutum*, *Brassica napus*, *Vitis vinifera*, *Daucus carota*, etc, and also in trees like *Populus trichocarpa*, *Cladrastis kentukea*, *Maackia amurensis*, *Sophora japonica*, *Sesbania rostrata*, *Prunus dulcis*, etc. In monocots, different types of extensins exist, containing two novel amino acid repeat motifs, ie, Thr-Pro-Lys-Pro-Thr-Hyp-Hyp-Thr-Tyr-Thr-Hyp-Ser-Hyp-Lys-Pro-Hyp- and Ala-Thr/Ser-Lys-Pro-Pro.³

Extensins are of many types, and generally extensin-1 and extensin-2 are present in a number of herbaceous plants and trees. Extensins 1–5 are in *A. thaliana* and, among these, 2–5 genes are strongly expressed during rehydration after dehydration.⁴ The present study considered extensin-1 and extensin-2 of herbaceous plants from *Fabaceae* and trees from different taxa in order to analyze accelerated evolution and also their role in abiotic stresses, like gravitational stress and tension wood formation, and biotic stress, like pathogenic resistance. The role of extensins involved in gravitational stress and tension wood formation was investigated through multiple sequence alignment analysis of the internal sequence repeats of extensin-1 and extensin-2, present in both herbs and trees. This study indicates that herbaceous plants may not be subjected to a significant amount of tensional forces, but are still subjected to gravitational forces, and that these proteins are associated with gravitational stress which, along with tensional stress, forms the two stress components of tension wood differentiation in trees. The role of extensins involved in pathogenic resistance was investigated by showing the existence of circular permutation in both extensin-1 and extensin-2 between plants of *Fabaceae* which was analyzed by Dot plots, and it was predicted that many partial circular permutations exist between the sequences of different length. In order to analyze for unequal crossover in extensin-2, a clustering study of the internal sequence repeats was done by constructing the phylogenetic tree and reconstructing the patterns, which shows the existence of unequal crossover in extensin-2. Furthermore, the present study shows that the accelerated evolution of extensin-2 and its role in pathogenic resistance arises from the existence of circular permutations and unequal crossover.

Methods

Sequences of the extensin-1 and extensin-2 family were identified from the Pfam database⁵ using Interproscan.⁶ From the Pfam database, the built-in phylogenetic trees of all of its sequences were observed as a reference. Based on the

relationship between sequences, the precursor sequences of extensins in the plants of *Fabaceae* and trees of different taxa were retrieved from the Swiss-Prot database.⁷

The individual internal sequence repeats of extensin-1 and extensin-2 were split into separate subsets of internal sequence repeats. The multiple sequence alignments were carried out for those subsets of internal sequence repeats using ClustalW.⁸ The consensus sequences are represented in the respective multiple sequence alignment using JalView.⁹

The partial circular permutations occurring in both extensin-1 and extensin-2 were analyzed using Dot plots¹⁰ from the CLC main workbench 5.0.2. For the unequal crossover analysis, the corresponding gene sequences of extensin-2 proteins were retrieved from the cross reference database EMBL¹¹ of Swiss-Prot. The analysis of extensin-2 gene structure was facilitated by the Xpro database.¹² A phylogenetic tree was constructed using the Neighbor-Joining method from MEGA4.¹³ Finally, the pattern of clusters from the phylogenetic tree was constructed and reconstructed manually to show the existence of unequal crossover.

Results and discussion

Internal sequence repeats of extensins

Sequences which were characterized from Pfam using the tool Interproscan as extensin-1 (PF02095) have 5–29 internal sequence repeats, and sequences which were categorized as extensin-2 (PF04554) have 1–11 internal sequence repeats. The selected extensin-1 and extensin-2 sequences of trees and herbs are listed in Tables 1 and 2, respectively, with their accession numbers, protein sequence lengths, and number of repeats. From Table 1, it was observed that the lengths of extensin-1 of fabaceous herbs such as *G. max*, *M. truncatula*, *P. sativum*, *M. sativa*, and *T. repens* vary in the range of 90–343 amino acid residues, and they have a varied number of internal sequence repeats in the range 5–29. In addition to fabaceous herbs, extensin-1 fragment sequences of trees from different taxa, such as *P. trichocarpa*, *S. rostrata*, *M. amurensis*, *C. kentukea* and *S. japonica*, were also included in this study to exhibit their role in tension wood formation. Their lengths vary in the range 182–330 amino acid residues, and the numbers of internal sequence repeats vary in the range 9–23. In total, 177 repeats of extensin-1 (98 from herbs and 79 from trees) were considered for the present study. Similarly, information about extensin-2 from Table 2 shows that the lengths of protein precursor sequences of *P. sativum*, *Vigna unguiculata*, *G. max*, and *P. vulgaris* vary in the range 144–580 amino acid residues and the numbers of internal sequence repeats vary in

Table 1 Sources of extensin-1 sequences of trees and herbs, along with their accession numbers, protein lengths, and number of repeats dealt with in this study

SI No.	Acc No. protein sequence	Protein/gene name	Species name	Protein sequence length	Internal repeats (n)
1	P15642	Repetitive proline-rich cell wall protein 3	<i>Glycine max</i> (H)	90	5
2	Q9FEW3	Putative repetitive proline-rich protein (ENOD11)	<i>Medicago truncatula</i> (H)	174	6
3	Q9SC42	Proline-rich protein (am4)	<i>Pisum sativum</i> (H)	214	16
4	Q40358	Repetitive proline-rich cell wall protein	<i>Medicago sativa</i> (H)	236	20
5	P08012	Repetitive proline-rich cell wall protein 1	<i>Glycine max</i> (H)	256	22
6	Q9MAW3	TrPRP2	<i>Trifolium repens</i> (H)	343	29
7	A9PE91	Predicted protein (POPTRDRAFT_566097)	<i>Populus trichocarpa</i> (T)	182	11
8	Q9FUR6	ENOD2 (fragment)	<i>Cladrastis kentukea</i> (T)	244	16
9	Q9FUR7	ENOD2 (fragment)	<i>Sophora japonica</i> (T)	269	9
10	Q9FUR5	ENOD2F (fragment)	<i>Maackia amurensis</i> (T)	308	20
11	Q41402	ENOD2 (fragment)	<i>Sesbania rostrata</i> (T)	330	23

Note: Total number of repeats: 98 (H) + 79 (T) = 177.

Abbreviations: Acc No., accession number; H, herb; T, tree; SI, serial.

the range 1–11. With these fabaceous herbs, trees of different taxa, such as *P. dulcis* and *P. trichocarpa*, were included and their lengths vary in the range 278–312 and numbers of internal sequence repeats vary in the range 2–4. In total, 41 repeats of extensin-2 (32 from herbs and nine from trees) were considered for the present study.

The amino acid compositions of extensin-1 and extensin-2 precursor sequences have been calculated, and the percentages of predominant amino acids are given in Tables 3 and 4, respectively, along with their isoelectric points and signal peptide ranges. From Tables 3 and 4, it is evident that both

extensin-1 and extensin-2 family sequences are basic in nature, and their isoelectric point values are approximately equal to 10.0 due to their high lysine content.³ Furthermore, both extensin-1 and extensin-2 are rich in proline. The study also reveals that the amino acids, such as lysine, proline, tyrosine, and valine, form more than 90% of the composition of many of the extensin-1 sequences, and lysine, proline, tyrosine, and serine form more than 90% of the composition of most of the extensin-2 sequences. The composition of valine is high in extensin-1 sequences compared with serine, and vice versa in extensin-2 sequences.

Table 2 Sources of extensin-2 sequences of trees and herbs, along with their accession numbers, protein lengths, and number of repeats dealt within this study

SI No.	Acc No. protein sequence	Protein/gene name	Species name	Protein sequence length	Internal repeats (n)
1	Q94ES5	Root nodule extensin	<i>Pisum sativum</i> (H)	144	1
2	Q76KW4	Hydroxyproline-rich glycoprotein-1	<i>Pisum sativum</i> (H)	152	1
3	Q94ES6	Root nodule extensin	<i>Pisum sativum</i> (H)	181	1
4	Q94ES9	Root nodule extensin	<i>Pisum sativum</i> (H)	183	1
5	Q94ES8	Root nodule extensin	<i>Pisum sativum</i> (H)	195	1
6	Q43687	Extensin-like protein	<i>Vigna unguiculata</i> (H)	242	2
7	Q9M6R7	Extensin	<i>Pisum sativum</i> (H)	327	6
8	Q39835	Extensin (SbHRGP3)	<i>Glycine max</i> (H)	432	8
9	Q09083	Hydroxyproline-rich glycoprotein	<i>Phaseolus vulgaris</i> (H)	580	11
10	Eugene3.00011698	Predicted protein	<i>Populus trichocarpa</i> (T)	256	3
11	Q40768	Extensin	<i>Prunus dulcis</i> (T)	278	2
12	B9NAX9	Predicted protein (POPTRDRAFT_587814)	<i>Populus trichocarpa</i> (T)	312	4

Note: Total number of repeats: 32 (H) + 9 (T) = 41.

Abbreviations: Acc No., accession number; H, herb; T, tree; SI, serial.

Table 3 Comparison of amino acid compositions of extensin-1 sequences, along with their isoelectric point values and signal peptide ranges

SI No.	Source protein Acc No.	Isoelectric point	Lys (%)	Pro (%)	Tyr (%)	Val (%)	Ser (%)	Signal peptide range
1	PI5642 (H)	9.94	17.8	26.7	13.3	10.0	2.2	1–24
2	Q9FEW3 (H)	10.72	15.5	31.0	3.4	2.3	5.7	1–25
3	Q9SC42 (H)	9.72	14.5	31.8	8.9	14.0	1.9	1–21
4	Q40358 (H)	9.83	16.9	36.4	12.3	19.5	0.8	1–22
5	P08012 (H)	9.87	17.6	35.9	14.5	15.2	2.3	1–26
6	Q9MAV3 (H)	9.92	17.2	37.3	11.1	21.3	0.9	1–22
7	A9PE91 (T)	9.74	18.1	33.5	9.9	9.9	1.1	1–22

Abbreviations: Acc No., accession number; H, herb; T, tree; Lys, lysine; Pro, proline; Tyr, tyrosine; Val, valine; Ser, serine; SI, serial.

Role of extensins in gravitational stress and tension wood formation

Although a number of functions have been attributed to extensin proteins,¹⁴ the most important is their association with almost all abiotic and biotic stresses.¹⁵ Through immunocytochemical methods, it was reported that extensins are associated with tension wood formation^{16,17} in the leaning stems and branches of trees, which are subjected to gravitational forces. Recently Lafarguette et al¹⁸ showed that arabinogalactan proteins are highly expressed in tension wood formation in *Populus*, but they had not made any observations on extensins and their relation to tension wood. More recently, Andersson-Gunnerås et al¹⁹ had also found that fasciclin-like arabinogalactan proteins are involved in tension wood differentiation. In addition, they had also found transcripts that encode extensin in the tension wood tissues of *Populus*. The present work has been undertaken from a bioinformatics perception to verify the involvement of extensins in tension wood formation, because tensile stresses are involved in this process.^{16,17} In addition, this work examines whether abiotic stress, like gravitational forces, are important in the

expression of extensins in tension wood, especially in light of the fact that many herbaceous taxa of plants do not have tension wood but they do have extensins.

Multiple sequence alignment of internal sequence repeats of extensins

The extensin-1 and extensin-2 repeat sequences were split into single internal sequence repeats separately and the multiple sequence alignments were constructed and the consensus sequences are represented in the respective multiple sequence alignments, which can be observed in Figures 1 and 2. In order to show the consensus residues of extensin-1 sequences from multiple sequence alignment, the repeats were manually arranged and all 177 repeats of both fabaceous herbs and trees from different taxa were named with accession numbers followed by the number of repeats and length of the repeats (eg, A9PE91_R1/1–11), as shown in Figure 1. One can also observe that several (177) small internal sequence repeats of extensin-1 are present with 10–11 amino acid residue lengths. Similarly, the repeats of extensin-2 were also manually arranged and all 41 repeats of

Table 4 Comparison of amino acid compositions of extensin-2 sequences, along with their isoelectric point values and signal peptide ranges

SI No.	Source protein Acc No.	Isoelectric point	Lys (%)	Pro (%)	Tyr (%)	Val (%)	Ser (%)	Signal peptide range
1	Q94E55 (H)	9.52	4.9	36.1	11.8	4.2	10.4	1–27
2	Q76KW4 (H)	9.64	5.3	34.9	9.9	4.6	13.2	1–30
3	Q94E56 (H)	9.44	4.4	38.1	11.0	5.0	10.5	1–27
4	Q94E59 (H)	9.44	4.4	38.8	10.9	4.9	10.4	1–27
5	Q94E58 (H)	9.54	5.1	37.9	10.8	5.6	10.8	1–27
6	Q43687 (H)	9.46	6.2	44.2	16.1	1.7	15.7	1–35
7	Q9M6R7 (H)	9.86	10.7	40.1	15.9	4.9	14.1	1–31
8	Q39835 (H)	9.36	10.0	46.3	18.3	2.3	11.6	1–23
9	Q09083 (H)	9.89	10.0	45.2	18.8	3.1	12.1	1–28
10	Eugene3.00011698 (T)	9.70	5.5	48.4	12.5	3.5	14.8	1–27
11	Q40768 (T)	9.81	10.1	40.6	12.2	4.7	14.7	1–25
12	B9NAX9 (T)	9.70	10.6	45.5	12.2	1.0	16.3	1–27

Abbreviations: Acc No., accession number; H, herb; T, tree; Lys, lysine; Pro, proline; Tyr, tyrosine; Val, valine; Ser, serine; SI, serial.

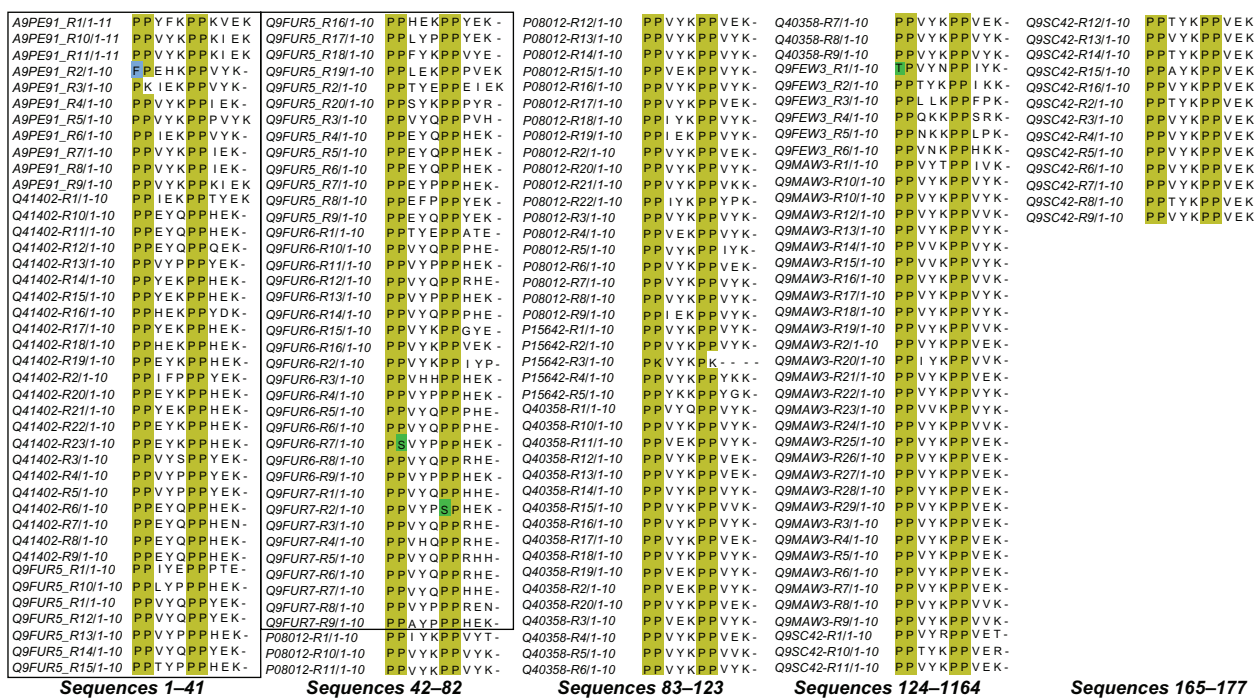


Figure 1 Multiple sequence alignment of the internal sequence repeats of extensin-I. Boxed sequences are of trees and the rest of the sequences are of herbaceous plants of *Fabaceae*. Extensin-I fragment sequences of trees (Q41402 *Sesbania rostrata*, Q9FUR5 *Maackia amurensis*, Q9FUR6 *Cladrastis kentukea*, Q9FUR7 *Sophora japonica*) were included in this alignment alone to exhibit the tension wood nature in trees and were not included in the unequal crossover and circular permutation studies.

both fabaceous herbs and trees of different taxa were named with accession numbers followed by the number of repeats and length of the repeats (eg, B9NAX9_R1/1-57) as shown in Figure 2. One can also observe that extensin-2 sequences have few (41) internal sequence repeats with 39–61 amino acid residue lengths.

From the multiple sequence alignment of extensin-1, it was observed that all the internal sequence repeats were aligned among trees (enclosed in box) in Figure 1 and also with the selected herbs. Similarly, from the multiple sequence alignment of extensin-2, it was observed that all the internal sequence repeats of trees were aligned with themselves (enclosed in box) in Figure 2 and with the repeats of other herbaceous plants. In extensin-1, proline residues are most often arranged in small contiguous clusters with the pattern PPxxxPPxxx (Figure 1) and four proline residues among them are highly conserved, as highlighted. Extensin-2 has small contiguous clusters with serine (Pro)₄, as shown in Figure 2, and one of the clusters is fully conserved among all the sequences, as highlighted. It shows that these four hydroxyproline residues in the PPxxxPPxxx motif in extensin-1 and serine (Pro)₄ in extensin-2 may play a very important role in abiotic stresses, like tension wood formation in the case of trees, and in gravitational stress, as in the case of both herbs and trees.

Generally, gravitational forces and tensile stresses are the two ultimate inductive causes of tension wood differentiation. Some investigators²⁰ argue that gravitational forces are the factors in tension wood differentiation, but do not support the operation of tensile stress in this process. However, another group of investigators argued for the operation of tensional forces in tension wood formation.¹⁷ Experimental studies done later^{17,21} proved that both gravitational force and imposed tensions are involved in the differentiation of tension wood. The clear existence of extensin-1 and extensin-2 family repeats in both herbaceous (showing response to gravitation, but not exhibiting tension wood) and arborescent (showing response to both gravitation and tension and exhibiting tension wood) taxa emphasizes that both extensin-1 and extensin-2 are involved in gravitation, but their role in tension wood formation needs to be viewed in the light of the following argument, already put forth by Krishnamurthy.¹⁷

This work cites experimental studies made by Fisher²¹ on *Terminalia catappa* and by Krishnamurthy¹⁷ and his group on *Caesalpinaceae* members, which had shown that both gravitational force and imposed tensions could cause the differentiation of tension wood, although the former stimulus was a dominant one when both factors operated simultaneously. The whole controversy was the result of two reasons, ie, a failure

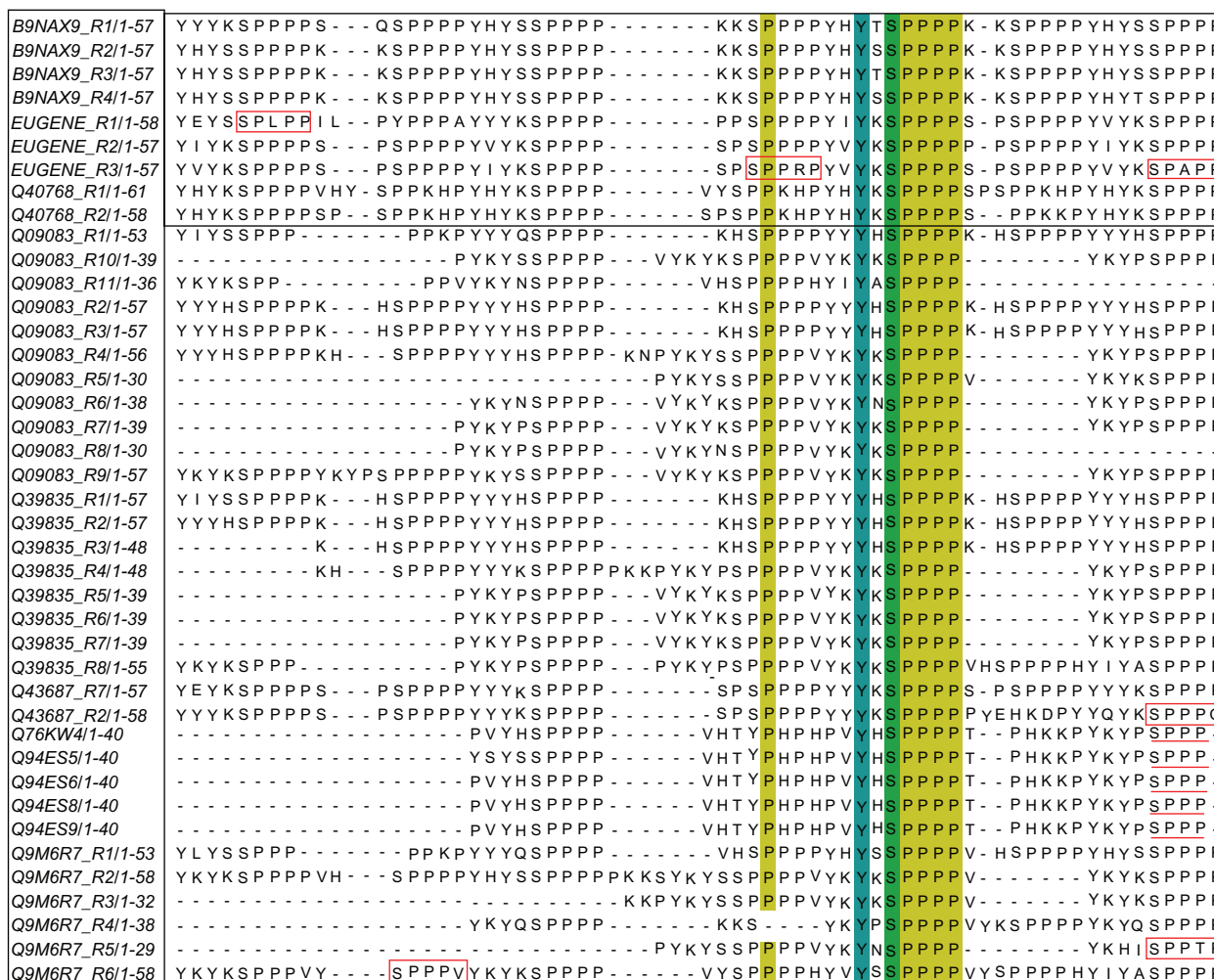


Figure 2 Multiple sequence alignment of the internal sequence repeats of extensin-2. Boxed sequences are of trees and the rest of the sequences are of herbaceous plants of *Fabaceae*.

to recognize the tension wood as possessing a syndrome of characteristics and laying emphasis only on the presence of gelatinous (G) fibers as a marker characteristic for tension wood, and a failure to understand that, although tension/tensional stress is a component of gravity, it can also result due to other factors not connected with gravity. When gravity acts on plant organs, such as horizontally disposed/leaning stems or roots, not only gravity-induced responses, but also tension-induced responses, are noticed in the organ concerned. It is the considered opinion of Krishnamurthy¹⁷ and his group that G fibers are the products of tension-induced stresses, while the tension wood syndrome, including the presence of G fibers, is the product of gravitational force which also imposes tensional stress. While the tension wood syndrome is possible only in cambium-derived products, G fibers can differentiate from noncambial meristems as well.

In light of the above, it is evident that the presence of extensin-1 and extensin-2 promotes or causes gravitational

response in both herbs and trees and the formation of tension wood due to the presence of cambium and gravitational force in trees. Perhaps the tensional stresses are associated with the tension wood and are also related to fasciclin-like arabinogalactan rich protein, as had been demonstrated in *Populus* by Lafarguette et al.¹⁸

Role of extensins in pathogenic resistance through circular permutation and accelerated evolution from unequal crossover

Apart from tension wood formation and environmental abiotic stress tolerance, extensin contributes to biotic stress, like wound healing, protecting against pathogenic attack, and plant defense.^{1,3} Immunochemical studies had also shown that extensins accumulate in the cell wall close to sites where microbial growth is restricted by the plant.¹ In the defense

mechanism, extensin may act as an impenetrable physical barrier or may immobilize the pathogen by binding to its surface. It probably results from positively charged extensin molecules interacting ionically with the negatively charged surfaces of certain pathogens.^{1,3} Because extensins play an important role in the defense mechanism, it was tempting to speculate that the underlying sequence rearrangements are part of a general scenario against pathogens. Barta et al²² analyzed the existence of unequal crossover in the circularly permuted Pin2 family of proteinase inhibitors of the plants of *Solanaceae* which has broad specificity against pathogens. In a similar way, the present study analyzed the existence of circular permutation and unequal crossover in extensin sequences as follows.

Circular permutation in extensin sequences and pathogenic resistance

Circular permutation in the protein is a rearrangement of the amino acid sequence, such that the original amino (N-) and carboxyl (C-) termini regions are interchanged. In other words, circular permutation in a protein means that the N-terminus of one protein is similar to the C-terminus of the other and vice versa.²³⁻²⁵ The permuted regions may cover a significant portion of both proteins, be within a partial region of the protein, or may be present in two proteins of very different lengths.²³ Studies of circular permutation are useful in many protein engineering and protein folding experiments.^{26,27}

Various researchers²⁷⁻²⁹ have reported the existence of circular permutation in many biological sequences and structures, including some carbohydrate-related enzymes and binding proteins, swaposins, transaldolases, FMN-binding

proteins, glutathione synthetases, methyltransferases, ferredoxins, and protease inhibitors. However, until now, the circular permutation is not observed in extensin sequences and structures. Therefore, the present study was undertaken to analyze circular permutation in extensin-1 and extensin-2 sequences from the herbaceous plants of *Fabaceae* and trees from different taxa.

A Dot plot was used to inspect circular permutation²³ from extensin sequences. The Dot plot results of both extensin-1 (P08012 versus Q9MAW3) and extensin-2 (Q39835 versus Q09083) are shown in Figures 3A and B, respectively. It was observed from Figure 3 that many diagonal lines originate from different rows and columns and they are the characteristic feature of circular permutation. It was also observed that many smaller diagonal lines in the Dot plot, which show many partial circular permutations exist in both the extensin-1 and extensin-2 sequences of different lengths. This preliminary study had predicted that circular permutation exists in both extensins and it also indicates that the circular permutation might have evolved to protect against pathogens (biotic stress). Previous studies^{25,29} also showed the existence of circular permutation in the type II family of proteinase inhibitors of potato to get protection from the pathogen.

Unequal crossover in extensin-2 and accelerated evolution

In order to exhibit the unequal crossover in the circularly permuted extensin-2, the evolutionary relationships were examined from their precursor protein sequences. As described in Figure 2, extensin-2 is rich in serine and hydroxyproline (SP₄) with the combination of other

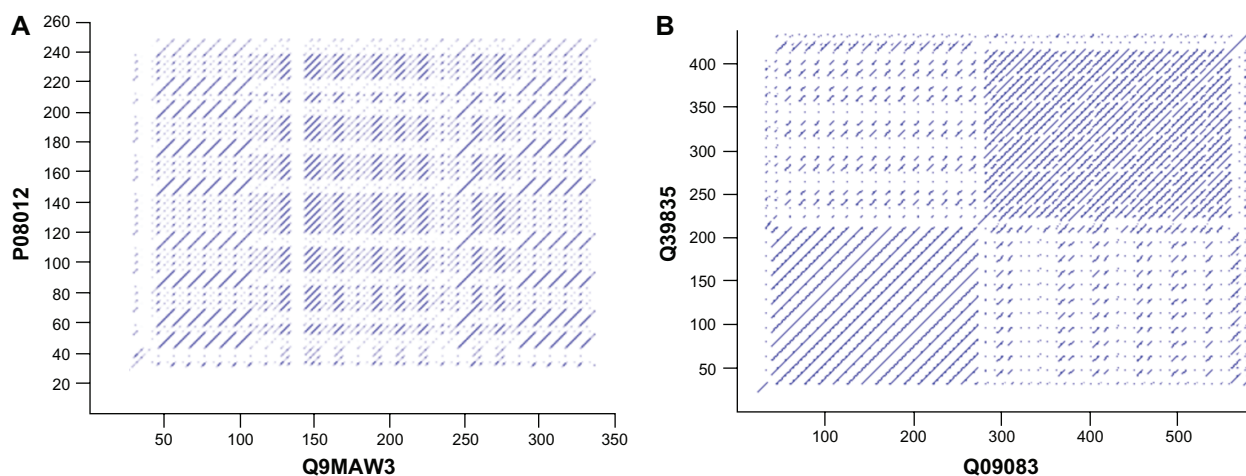


Figure 3 Dot plot view of circular permutation of extensin-1 (**A** P08012 versus Q9MAW3) and extensin-2 (**B** Q39835 versus Q09083) protein pairs.

amino acids, such as tyrosine, histidine, and lysine, in high percentage, as given in Table 4. The length of extensin-2 varies from plant to plant among *Fabaceae*, although there are some notable similarities among them. In order to retrieve gene and protein sequences of extensin-2, the Pfam database was used as a reference, from where hydroxyproline-rich glycoprotein sequences containing the extensin-2 family (PF090554) were obtained. It has 352 sequences, and many of them have come from plants of *Fabaceae*, others from plants including *Arabidopsis*, *Solanum*, *Nicotiana*, *Catharanthus*, *Daucus*, *Brassica*, etc. Among them, the precursor sequences of related herbaceous *Fabaceae* and tree sequences (which have already been included for tension wood study) were considered for this analysis. Proteins and their corresponding gene sequences were retrieved from Swiss-Prot and EMBL databases, respectively, and their accession numbers are B9NAX9 (XM_002332377), Q40768 (X65718), Q09083 (U18791), Q39835 (U44838), Q43687 (X86029), Q76 KW4 (AB087834), Q94ES5 (AE397030), Q94ES6 (AF397030), Q94ES8 (AF397027), Q94ES9 (AF397026), and Q9M6R7 (AF155232). Because the single-repeat domain of the nonfabaceous herb *Nicotiana plumbaginifolia* (Q40402 [M34371]), which was identified from the Pfam database, is related to the fabaceous sequences, it was also included in the analysis.

Extensin-2 sequences of the selected plant species have a similar gene structure in that they have a single exon which codes a signal peptide followed by internal sequence repeats and there is no intron. We performed a systematic sequence comparison on the repeats through multiple sequence alignment and phylogenetic tree construction. For better identification, the repeats were numbered as EP1-accession number-R1 through EP11-accession number-R11 (eg, EP7-Q39835-R7, where EP7 denotes an extensin precursor that has seven internal sequence repeats, Q39835 is the accession number and R7 denotes the seventh internal sequence repeat), starting from the N terminus of the precursors. In the multiple sequence alignment of the internal sequence repeats of extensin-2, one serine, four proline (SP₄), and one tyrosine residue are completely conserved (Figure 2). Besides these, also in other positions, serine, proline, tyrosine, and lysine residues are highly conserved which can be observed in Figure 2. It was also observed that in each repeat, variable numbers of the SP₄ motif are present and are in the range 3–7 (compare with Figure 2). Few repeats, such as Eugene_R1, Eugene_R3 (in two positions), Q43687_R2,

Q9M6R7_R5, and Q9M6R7_R6, have a single residue change in the SP₄ motif (SPLPP, SPPRP, SPAPP, SPPPQ, SPPTP, and SPPPV, respectively), and the repeats of Q76 KW4, Q94ES5, Q94ES6, Q94ES8, and Q94ES9 have an SP₃ motif instead of SP₄, which are enclosed in the box and underlined, respectively, in Figure 2. This observation shows that the presence of the SP₃ motif and a single residue change in SP₄ motifs (SPLPP, SPPRP, SPAPP, SPPPQ, SPPTP, and SPPPV) indicates that they are transforming during the course of evolution to become a motif like SP₄ to tolerate abiotic stresses, like gravitational stress and tension wood formation in the case of trees, and biotic stress, like pathogenic resistance.

To investigate the accelerated evolution in the extensin-2 family, the phylogenetic tree was inferred from 42 internal sequence repeats of extensin-2 of eight different taxa using MEGA4¹³ with the Neighbor-Joining method³⁰ choosing default parameters. The bootstrap consensus tree inferred from 500 replicates^{31,32} is taken to represent the evolutionary relationships of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches.^{31,32} The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method³³ and are in the units of the number of amino acid substitutions per site by MEGA4. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). In total, there were 19 positions in the final dataset. The obtained bootstrap consensus tree with clusters is shown in Figure 4 with their corresponding percentage of replicate trees encircled. We considered the above tree for further unequal crossover analysis.

In the phylogenetic tree, the taxa are clustered into three clades based on the number of conserved repeats. In cluster 1, the entire single repeats of both *P. sativum* and *N. plumbaginifolia* are clustered together. In cluster 2, R1–R4 repeats of *P. trichocarpa*, *P. vulgaris*, *G. max*, *P. dulcis*, and *V. unguiculata* are clustered together, and in cluster 3, R6–R11 repeats of *P. sativum*, *G. max*, and *P. vulgaris* are clustered together. From the above observation on the conserved repeats of phylogenetic tree, the results are discussed as follows. The tree sequences (R1–R4 of EP4-B9NAX9 and R1–R3 of EP3-Eugene) are clustered with that of the plants of *Fabaceae* in cluster 2 and it may be attributed to the fact

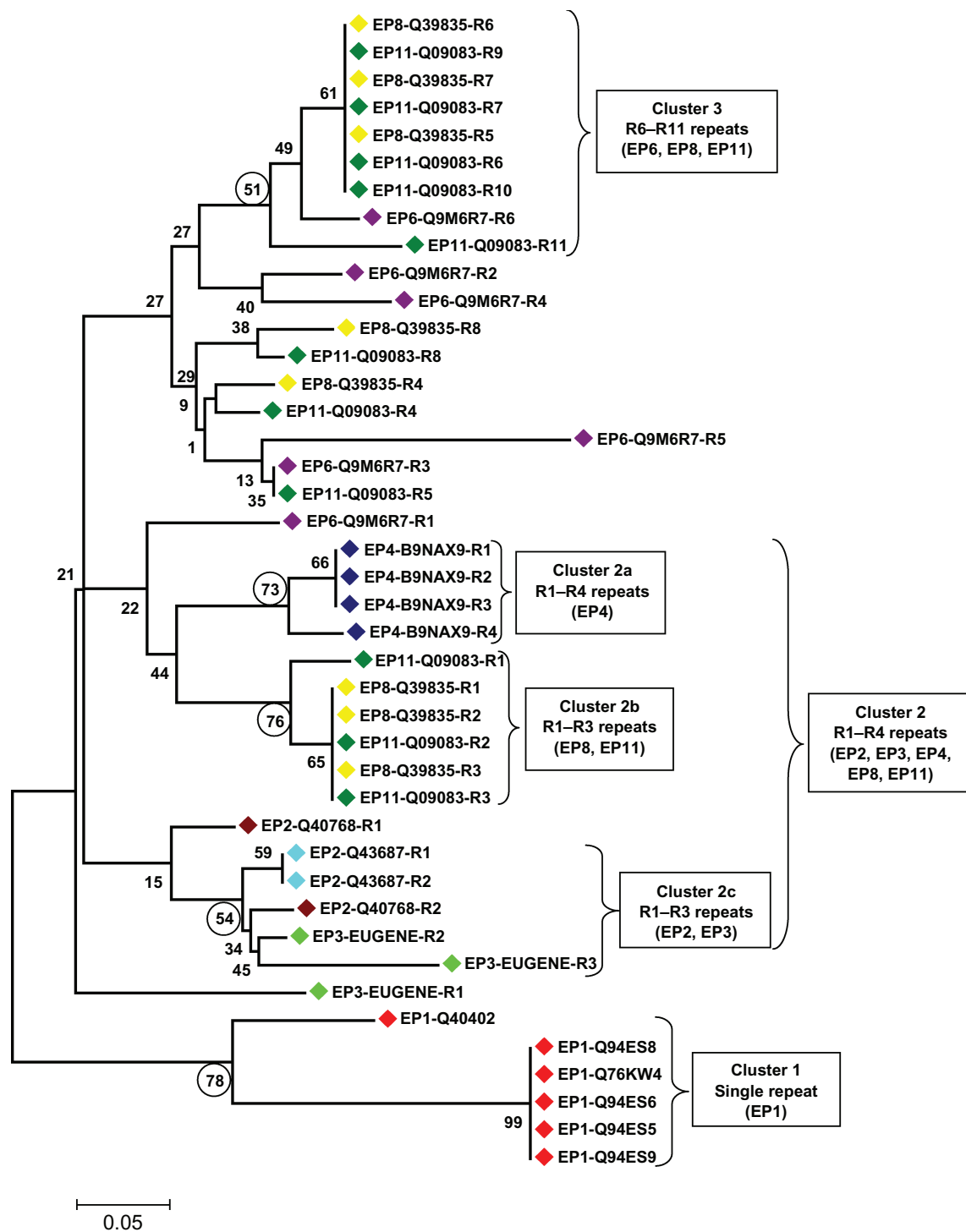


Figure 4 Evolutionary relationships of 42 internal sequence repeats of extensin-2 of eight different taxa were inferred using the Neighbor-Joining method (Bootstrap consensus tree inferred from 500 replicates) using MEGA4. Clusters were made based on the group of repeat units with more than 50% bootstrap replicates (encircled) between sequences. Sequence of the nonfabaceous plant *Nicotiana plumbaginifolia* (Q40402) was included for this analysis alone.

that extensin-2 is involved in pathogenic resistance (biotic stress) in fabaceous herbs and involved in tensile stress (abiotic stress) in the case of trees. Genes of *P. sativum* (Q94ES5, Q94ES6, Q94ES8, Q94ES9, and Q76 KW4) and *N. plumbaginifolia* (Q40402) encode single extensin-2 repeat

units, in which there is no repeat expansion. These genes are likely to be an ancestral form where the initial multiplication occurred in *Fabaceae*. The single domain repeat of the nonfabaceous plant, *N. plumbaginifolia* (Q40402), can also be considered as its ancestor.

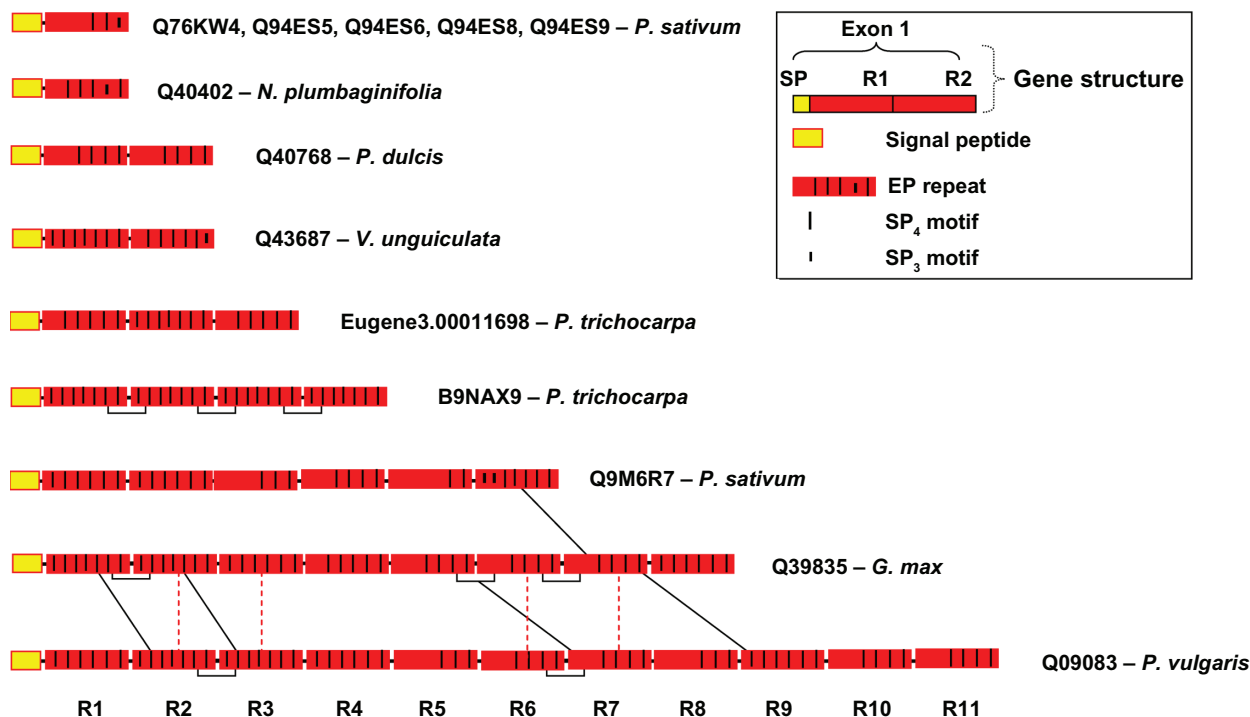


Figure 5 The pattern of internal sequence repeats of extensin-2 of eight different taxa. From the three clusters of phylogenetic tree, the pattern was constructed manually to show the presence of identical repeats within and between (adjacent) sequences, marked with lines in order to show the unequal crossover.

From the clusters of phylogenetic tree, the pattern was constructed manually to show the identical repeats among sequences. The identical repeats within the sequence and identical and similar repeats (which are grouped in clusters 2b and 3) between sequences are shown through lines in Figure 5. From the above constructed model, the general architecture of the extensin-2 family repeats show that the domain expansion occurs in the plants of *Fabaceae*. Also the pattern of repeat identities was reconstructed from Figure 5, by a series of putative unequal crossover within exon for extensin-2, and two types of unequal crossover events (adjacent, $R_1 \times R_{i+1}$ and nonadjacent, $R_1 \times R_{i+2}$) were observed, which are shown in Figure 6. The above mentioned patterns present in between the internal sequence repeats as shown in Figure 5 along with their percentage of sequence identities in the global and local pairwise sequence alignments are given (within brackets) as follows. The pattern $R_1 \times R_{i+1}$ is shown in three repeats EP6-Q9M6R7-R6 versus EP8-Q39835-R7 (global 55.2% and local 70.7%), EP8-Q39835-R1 versus EP11-Q09083-R2 (global and local 96.5%) and EP8-Q39835-R2 versus EP11-Q09083-R3 (global and local 100%). Similarly, the $R_1 \times R_{i+2}$ pattern is shown in two repeats EP8-Q39835-R5 versus EP-Q09083-R7 (global and local 100%) and EP8-Q39835-R7 versus EP11-Q09083-R9 (global 66.7% and local 97.4%). In addition to these two patterns, another pattern $R_1 \times R_{i+3}$

also exists, which is shown in two repeats EP8-Q39835-R6 versus EP11-Q09083-R9 (global 66.7% and local 97.4%) and EP8-Q39835-R7 versus EP11-Q09083-R10 (global and local 97.4%), but this pattern is not included in Figures 5 and 6. Therefore, the above analysis clearly indicates that the unequal crossover exists in extensin-2 of different plants of *Fabaceae*. This leads us to the logical conclusion that the accelerated evolution of this gene might have been triggered by an initial repeat multiplication that occurred in ancestral *Fabaceae*. It thus appears that, in response to pathogenic attack, plants resort to extensins similar to those used to fight retroviral infections in humans, as in the proteinase inhibitors of the *Solanaceae* family.²² A similar analysis was performed

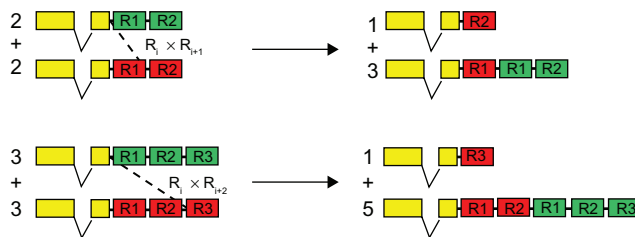


Figure 6 Some of the potential unequal crossover events that explain the emergence of sequence identity patterns of the extensin-2 sequences. The patterns are colored in red and green. The types of unequal crossover events involving either adjacent ($R_1 \times R_{i+1}$) or nonadjacent ($R_1 \times R_{i+2}$) internal sequence repeats are shown by dashed lines.

with extensin-1 and several internal sequence repeats with the pattern PPxxxPPxxx were observed, which indicate that the presence of unequal crossover. Because several internal sequence repeats exist between and within sequences, the unequal crossover nature of extensin-1 was not clearly indicated as in the case of extensin-2.

Conclusion

The present computational internal sequence repeats analysis provides a systematic analysis of the extensin-1 and extensin-2 family in a significant dataset by considering previous studies^{17,18,22,23} explaining the accelerated evolution and the role of these proteins in abiotic stresses, like gravitational stress in both herbs and trees and tension wood formation in trees, and biotic stress, like pathogenic resistance.

Multiple sequence alignment analysis of extensin-1 and extensin-2 sequences shows that the internal sequence repeats of both the extensins are aligned among trees and also with the selected fabaceous herbs. The proline residues are particularly highly conserved in extensin-1, and the SP₄ motif is fully conserved in the case of extensin-2. It shows that those conserved residues, ie, proline and serine (Pro)₄, may play a very important role in abiotic stresses, like tension wood formation in trees and gravitational stress in the case of both herbs and trees.

Apart from abiotic stress tolerance, extensins also contribute to biotic stress like plant defense and pathogenic resistance. The Dot plots show many diagonal lines originating from different rows and columns of extensin-1 and extensin-2, and they explain that partial circular permutations exist in both the extensin sequences. These results indicate that the circular permutation is for the tolerance of biotic stress, like pathogenic resistance.

Similarly, in order to exhibit unequal crossover, the internal sequence repeats of extensin-2 family were considered for phylogenetic tree construction. From the clusters of the phylogenetic tree, the pattern was constructed and reconstructed manually and showed the existence of unequal crossover in extensin-2. From this analysis, it has been found that the single domain repeats of fabaceous herbs and the nonfabaceous herb, *N. plumbaginifolia*, may be considered as its ancestor. It was also predicted that the accelerated evolution of this extensin gene might have been triggered by an initial repeat multiplication that occurred in ancestral *Fabaceae* and thus appears in response to biotic stress, like pathogenic resistance.

Disclosure

The authors report no conflicts of interest in this work.

References

- Cassab GI. Plant cell wall proteins. *Ann Rev Plant Physiol Plant Mol Biol.* 1998;49:281–309.
- Cassab GI, Varner JE. Cell wall proteins. *Ann Rev Plant Physiol Plant Mol Biol.* 1988;39:321–353.
- Showalter AM. Structure and function of plant cell wall proteins. *Plant Cell.* 1993;5:9–23.
- Yoshida Y, Aoki C, Iuchi S, et al. Characterization of four extensin genes in *Arabidopsis thaliana* by differential gene expression under stress and non-stress conditions. *DNA Res.* 2001;8:115–122.
- Finn RD, Tate J, Mistry J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2008;36:D281–D288.
- Labarga A, Valentin F, Andersson M, Lopez R. Web services at the European Bioinformatics Institute. *Nucleic Acids Res.* 2007;35:6–11.
- The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2007;35:D193–D197.
- Larkin MA, Blackshields G, Brown NP, et al. Clustal W and ClustalX version 2. *Bioinformatics.* 2007;23:2947–2948.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25:1189–1191.
- Maizel JV Jr, Lenk RP. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc Natl Acad Sci U S A.* 1981;78:7665–7669.
- Kanz C, Aldebert P, Althorpe N, et al. The EMBL nucleotide sequence database. *Nucleic Acids Res.* 2005;1:D29–D33.
- Gopalan V, Tan TW, Lee BT, Ranganathan S. Xpro: Database of eukaryotic protein-encoding genes. *Nucleic Acids Res.* 2004;1:D59–D63.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007;24:1596–1599.
- Krishnamurthy KV. *Methods in Cell Wall Cytochemistry.* Washington, DC: CRC Press; 1999.
- Showalter AM, Rumeau D. Molecular biology of plant cell wall hydroxyproline-rich glycoproteins. In: Adair WS, Mecham RP, editors. *Organization and Assembly of Plant and Animal Extracellular Matrix.* New York, NY: Academic Press; 1990.
- Hariharan Y, Krishnamurthy KV. A cytochemical study of cambium and its xylary derivatives on the normal and tension wood sides of the stems of *Prosopis juliflora* (SW). *Beitr Biol Pflanzen.* 1995;69:459–472. German.
- Krishnamurthy KV. In pursuit of understanding the tension wood riddle. *J Indian Bot Soc.* 2007;86:14–26.
- Lafarguette F, Leple JC, Dejardin A, et al. Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol.* 2004;164:107–121.
- Andersson-Gunnerås S, Mellerowicz EJ, Love J, et al. Biosynthesis of cellulose-enriched tension wood in *Populus*: Global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J.* 2006;45:144–165.
- Wilson BF. The development of growth strains and stresses in reaction wood. In: Barnett JR, editor. *Xylem Cell Development.* Kent: Castle House Publications; 1981.
- Fisher JB. Induction of reaction wood in *Terminalia* (Combretaceae): Roles of gravity and stress. *Ann Bot.* 1985;55:237–248.
- Barta E, Pintar A, Pongar S. Repeats with variations: Accelerated evolution of the Pin2 family of proteinase inhibitors. *Trends Genet.* 2002;18:600–603.
- Uliel S, Fliess A, Unger R. Naturally occurring circular permutations in proteins. *Protein Eng.* 2001;14:533–542.
- Weiner J III, Bornberg-Bauer E. Evolution of circular permutations in multi-domain proteins. *Mol Biol Evol.* 2006;23:734–743.
- Kong L, Ranganathan S. Tandem duplication, circular permutation, molecular adaptation: How *Solanaceae* resist pests via inhibitors. *BMC Bioinformatics.* 2008;9 Suppl 1:S1–S22.

26. Wei-Cheng L, Ping-Chiang L. CPSARST: An efficient circular permutation search tool applied to the detection of novel protein structural relationships. *Genome Biol.* 2008;9:R11.1–R11.14.
27. Wei-Cheng L, Chi-Ching L, Che-Yu L, Ping-Chiang L. CPDB: A database of circular permutation in proteins. *Nucleic Acids Res.* 2009;37:D328–D332.
28. Lindqvist Y, Schneider G. Circular permutations of natural protein sequences: Structural evidence. *Curr Opin Struct Biol.* 1997;7:422–427.
29. Antcheva N, Pintar A, Patthy A, et al. Proteins of circularly permuted sequence present within the same organism: The major serine proteinase inhibitor from *Capsicum annum* seeds. *Protein Sci.* 2001;10:2280–2290.
30. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4:406–425.
31. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 1985;39:783–791.
32. Efron B, Tibshirani RJ. *Introduction to the Bootstrap.* New York, NY: Chapman and Hall/CRC; 1993.
33. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. *Evolving Genes and Proteins.* New York, NY: Academic Press; 1965.

Open Access Bioinformatics

Dovepress

Publish your work in this journal

Open Access Bioinformatics is an international, peer-reviewed, open access journal publishing original research, reports, reviews and commentaries on all areas of bioinformatics. The manuscript management system is completely online and includes a very quick and fair

peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/open-access-bioinformatics-journal>