J Mol Evol (1997) 44:614–624

JOURNAL OF MOLECULAR EVOLUTION

Structural and Evolutionary Relationships Among Chitinases of Flowering Plants

Francine Hamel,[1] Rodolphe Boivin,[2] Colette Tremblay,[1] Guy Bellemare[1]

[1]Département de biochimie, Faculté des sciences et de génie, Pavillon Charles-Eugène-Marchand, Université Laval, Québec Qc, Canada G1K 7P4
[2]Centre de recherche en biologie forestière, Pavillon Charles-Eugène-Marchand, Université Laval, Québec Qc, Canada G1K 7P4

Received: 5 July 1996 / Accepted: 9 January 1997

Abstract. The analysis of nuclear-encoded chitinase sequences from various angiosperms has allowed the categorization of the chitinases into discrete classes. Nucleotide sequences of their catalytic domains were compared in this study to investigate the evolutionary relationships between chitinase classes. The functionally distinct class III chitinases appear to be more closely related to fungal enzymes involved in morphogenesis than to other plant chitinases. The ordering of other plant chitinases into additional classes mainly relied on the presence of auxiliary domains—namely, a chitin-binding domain and a carboxy-terminal extension—flanking the main catalytic domain. The results of our phylogenetic analyses showed that classes I and IV form discrete and well-supported monophyletic groups derived from a common ancestral sequence that predates the divergence of dicots and monocots. In contrast, other sequences included in classes I* and II, lacking one or both types of auxiliary domains, were nested within class I sequences, indicating that they have a polyphyletic origin. According to phylogenetic analyses and the calculation of evolutionary rates, these chitinases probably arose from different class I lineages by relatively recent deletion events. The occurrence of such evolutionary trends in cultivated plants and their potential involvement in host–pathogen interactions are discussed.

Key words: Endochitinase evolution — Class I, II, III, and IV — Nuclear genes — PR proteins — Monocots — Dicots — Angiosperms

Correspondence to: G. Bellemare; e-mail Guy.Bellemare@rsvs.ulaval.ca

Introduction

The mechanisms evolved by plants to protect themselves against pathogen invasion include the synthesis of pathogenesis-related (PR) proteins. Among the PR proteins synthesized in response to fungal infection or treatment with fungal elicitors, several have been identified as chitinases (Legrand et al. 1987). Endochitinases (EC 3.2.1.14) hydrolyze chitin, a homopolymer of β-1,4-*N*-acetyl-D-glucosamine. Chitin constitutes an important component of fungal cell walls and arthropod or nematode exoskeletons (Boller 1988). Chitinases are present in all plants analyzed to date and many have been shown to inhibit fungal growth either in vitro (Schlumbaum et al. 1986; Leah et al. 1991) or when expressed in transgenic plants (Broglie et al. 1991; Jach et al. 1995). Chitinases seem to be encoded by a relatively small number of genes in angiosperms since only a limited number of distinct sequences have been recovered, even from plant genomes subjected to extensive molecular survey (Samac et al. 1990; Van Buuren et al. 1992; Danhash et al. 1993; Hamel and Bellemare 1993; Vogelsang and Barz 1993).

Following comparisons of chitinase sequences from various plant species, separation into distinct classes has been suggested (Shinshi et al. 1990; Meins et al. 1992; Collinge et al. 1993; Melchers et al. 1994; Beintema 1994; Araki and Torikata 1995). Classes I, II, and IV

share a homologous main catalytic domain in addition to the signal peptide found in all chitinases. Class I chitinases, found in both monocots and dicots, possess a cysteine-rich domain involved in chitin binding in their amino-terminal region (Iseli et al. 1993). Most class I chitinases are targeted to the vacuole by means of an essential carboxy-terminal signal (Neuhaus et al. 1991). Class II chitinases, found mainly in dicots, lack the cysteine-rich domain (Shinshi et al. 1990). The carboxy-terminal vacuolar targeting signal is also absent, so these chitinases are secreted to the apoplast. Class IV comprises a group of extracellular chitinases containing a cysteine-rich domain and sharing two characteristic deletions in the main catalytic domain. This class has been identified mainly in dicots (Margis-Pinheiro et al. 1991; Mikkelsen et al. 1992; Rasmussen et al. 1992; Collinge et al. 1993). Class III includes bifunctional lysozyme/chitinase enzymes (Jekel et al. 1991) with no sequence similarity to plant chitinases from other classes (Meins et al. 1992). Class III chitinases were identified in various dicots (Métraux et al. 1989; Samac et al. 1990; Lawton et al. 1992; Ishige et al. 1993; Nielsen et al. 1993; Vogelsang and Barz 1993) and in the monocot barley (Kragh et al. 1993). Thirteen class III–like expressed sequences tags (EST) were also found in rice.[1] Other proteins with endochitinase activity but unrelated to other plant chitinases have been isolated from tobacco: They share some similarity with bacterial exochitinases and they have been designated as class V (Melchers et al. 1994).

The crystal structure of a barley chitinase has been determined by Hart et al. (1993; 1995). Several comparative studies, using these data as a starting point, concluded that class I, II, and IV chitinases share a very well conserved structure in their catalytic domain (Beintema 1994), including the so-called lysozyme catalytic fold (Holm and Sander 1994; Monzingo et al. 1996). This feature marks them as members of the lysozyme superfamily, likely to have arisen by divergent evolution (Monzingo et al. 1996).

A few studies have been published on the molecular evolution of chitinases in flowering plants (Davis et al. 1991; Hamel 1994; Araki and Torikata 1995). In the first paper, Davis et al. analyzed the relationships among 13 chitinase catalytic domains from five dicot genera. Phylogenetic reconstruction in Araki and Torikata (1995) involved distance analysis of 35 amino acid sequences from class I, II, and IV chitinases in both dicots and monocots. In this report, which expands the work of Hamel (1994), a different and more extensive survey is conducted using mainly distance analyses of 54 nucleotide sequences from class I, II, IV, as well as class III chitinase genes from 23 angiosperm taxa, providing new insights into the origin and evolution of chitinases.

Materials and Methods

Sequences. Nucleotide sequences of various chitinase genes were obtained from the GenBank database release 91 (see Table 1 and Fig. 2). Only complete sequences, up to the 5′ end of the coding region, were selected for sequence comparisons. Percentage identity scores were based on sequences aligned with the Bestfit and Pileup programs from the UWGCG Sequence Analysis Software Package (Devereux et al. 1984). In addition, alignments were refined by hand by adjusting gap positions to enhance overall sequence identity. The conserved domain used for analyses of class III and related sequences corresponds to codons 30 to 290 of the *Athchia* gene from *Arabidopsis thaliana* (GenBank Acc. No. M34107; Samac et al. 1990). For the analyses of class I, II, and IV sequences, we used a data matrix of 777 positions (containing 198 to 246 codons) comprising the complete main catalytic domain. At the 5′ end, sequences upstream of codon 71 in *Brassica napus BnCh25* (GenBank Acc. No. M95835; Hamel and Bellemare 1993) were truncated to remove sequences coding for the signal peptide, the cysteine-rich domain, and the hinge region found in class I (see structure outline in Fig. 1). Sequences were truncated at the 3′ end after codon 312 in *BnCh25* because little conservation exists beyond this position among all chitinase genes.

Phylogenetic Analyses. Calculation of evolutionary rates and the neighbor-joining method of phylogenetic tree construction (Saitou and Nei 1987) were conducted using the MEGA analysis platform (Kumar et al. 1993) with the pairwise-deletion option. Proportions of synonymous (P_S) and nonsynonymous (P_A) substitutions per synonymous and nonsynonymous site, respectively, were calculated from the method of Nei and Gojobori (1986). Pairwise comparisons with the corresponding corrected rates (K_S and K_A) were not possible, since computation of K_S mostly resulted in invalid distance values. For the neighbor-joining analyses, rates of nucleotide substitution per site were estimated by the two-parameter method of Kimura (1980). Other substitution rate models available in the MEGA software—namely, Tamura and Tamura-Nei—gave identical topologies and nearly the same bootstrap values. Confirmatory work also involved parsimony analyses of nucleotide sequences using PAUP v. 3.1.1 (Swofford 1993). In phylogenetic tree reconstructions, the confidence level that could be assigned to the various nodes was determined by 500 bootstrap replications.

Results

Classification of Chitinase Sequences

Several nucleotide sequences coding for plant chitinases were classified according mainly to the structural features of their deduced proteins (outlined in Fig. 1). Twenty-two sequences were included in class I (Table 1): They all encoded a cysteine-rich domain (CRD) and a carboxy-terminal extension (CTE), and they showed an average identity score of 66% within their catalytic domain. Six sequences isolated mostly from monocots and encoding a CRD were grouped with class I although they lacked CTE, because their mean identity level was higher (65%) with class I than with any other group.[2] They were

[1]GenBank accession numbers: D46335, D41205, D21969, D41935, D22071, D22311, D22287, D15341, D22372, D40380, D22225, D22210, and D21990

[2]Among these figures the barley enzyme whose crystal structure have been determined by Hart et al. (1993; 1995)

Table 1. Classification of plant taxa and chitinase sequences analyzed in this study

Class	Subclass	Family	Species	Common name	Class of chitinase[a]	GenBank Accession No.
Magnoliopsida						
	Asteridae					
		Caprifoliaceae	*Sambucus nigra*	European elder	IV	(a) Z46948, (b) Z46950
		Solanaceae	*Lycopersicon esculentum*	Tomato	I	Z15140
					II	(a) Z15141, (b) Z15139
			Nicotiana tabacum	Tobacco	I	(a) X16938, (b) X64519, (c) X64518
					II	(a) M29868, (b) M29869
					III	(a) Z11563, (b) Z11564
			Petunia × hybrida	Petunia	II	X51427
			Solanum tuberosum	Potato	I	(a) X15494, (b) X14133
					II	X67693
	Caryophyllidae					
		Chenopodiaceae	*Beta vulgaris*	Beet	I	X79301
					III	S66038
					IV	(a) A23392, (b) L25826
	Dilleniidae					
		Brassicaceae	*Arabidopsis thaliana*	Thale cress	I	M38240
					III	M34107
			Brassica napus	Rape	I	M95835
					IV	X61488
		Cucurbitaceae	*Cucumis sativus*	Cucumber	III	(a), (b), (c) M84214
		Salicaceae	*Populus trichocarpa × P. deltoides*	Poplar	I	(a) M25336, (b) M25337
		Sterculiaceae	*Theobroma cacao*	Cacao	I	U30324
	Hamamelidae					
		Ulmaceae	*Ulmus americana*	American elm	I	L22032
	Rosidae					
		Fabaceae	*Cicer arietinum*	Chick pea	III	X70660
			Phaseolus vulgaris	Kidney bean	I	(a) M13968, (b) S43926
					IV	X57187
			Pisum sativum	Pea	I*	(a) X63899
					I	(b) L37876
			Psophocarpus tetragonolobus	Winged bean	III	D49953
			Vigna angularis	Adzuki bean	III	D11335
			Vigna unguiculata	Cowpea	I	X88800
					III	X88801
		Vitaceae	*Vitis vinifera*	Grapevine	I	Z54234
Liliopsida						
	Commelinidae					
		Poaceae	*Hordeum vulgare*	Barley	I*	(a) U02287
					I	(b) L34211
					II	M62904
			Oryza sativa	Rice	I*	(a) X54367, (b) X56063
					I	(c) L37289, (d) X56787
			Triticum aestivum	Wheat	I*	X76041
			Zea mays	Maize	I	(a) L16798
					I*	(b) L00973
					IV	(a) M84164, (b) M84165

[a] Special class designations: I*, class I lacking a carboxy-terminal extension

designated I* in Table 1. Class II designation was used for chitinase genes lacking coding sequences both for CRD and CTE. Nucleotide sequences in the catalytic domain showed from 56% to 94% identity within class II. This class included sequences from *Lycopersicon esculentum*, *Nicotiana tabacum*, and *Petunia hybrida* with a 14-codon deletion in the catalytic domain not found in class I chitinases, and sequences from *Solanum tuberosum* and *Hordeum vulgare* that do not possess this deletion (Fig. 1). Eight of the remaining sequences from five angiosperm subclasses were characterized by the presence of a CRD and the absence of a CTE in the deduced proteins; they were classified within class IV (Table 1). All these sequences shared two deletions of 13 and 19 codons, respectively, in their catalytic domain, while seven codons were missing at the 3′ end of the sequences (Fig. 1). The class IV sequences shared an average identity level of 63% among each other, but only

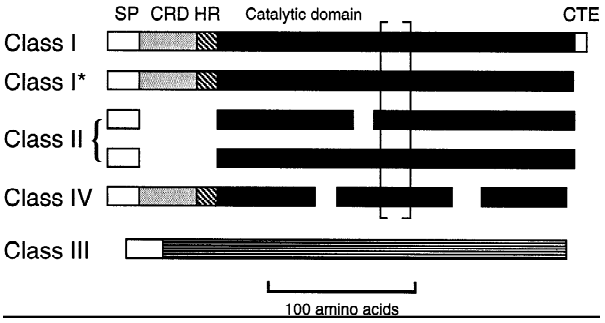

Fig. 1. Schematic representation of the structural differences between chitinase classes in angiosperms. Gaps found in most class II and in class IV chitinases, when compared to class I, are indicated as such. The signal peptide (*SP*) and the carboxy-terminal extension (*CTE*) are represented by *open boxes*, while the cysteine-rich domain (*CRD*) and the hinge region (*HR*) are identified by *shadowed and hatched boxes*, respectively. *Filled boxes* represent the catalytic domain, used in this work for phylogenetic analysis of classes I, II, and IV, while *horizontally hatched box* identifies the divergent class III catalytic domain. In one region of the catalytic domain, *brackets* delineate a previously identified catalytic site (Verburg et al. 1993).

of 53% and 51% with class I and class II sequences, respectively. In spite of their distinctive features, class IV sequences appeared clearly related, albeit distantly, to class I and class II sequences. On the other hand, no significant relationship, even in the conserved catalytic domain, was observed by comparing the remaining sequences, grouped into class III, with class I, class II, or class IV sequences. This result prompted us to analyze the class III sequences separately. No class V sequences were analyzed in this study.

Relationships Among Class III Plant Chitinases and Comparisons with Fungal Chitinases

Screening of the sequence databases revealed that class III plant chitinases had higher sequence identity with yeast (*Saccharomyces cerevisiae* and *Candida albicans*) and zygomycetes (*Rhizopus niveus* and *R. oligosporus*) chitinase sequences than with plant chitinases belonging to other classes. Class III chitinases also shared significant identity and overall structural similarity with concanavalin B, a plant protein stored in *Canavalia ensiformis* seed cotyledons (GenBank Acc. No. M83426). Within the domain conserved between plant class III and fungal chitinases (Fig. 2), amino acid identity scores ranged from 48% to 74% within the yeast-zygomycetes group and from 29% to 43% between plant and fungal sequences. In the 285-amino-acid alignment presented in Fig. 2, some regions were well conserved among all sequences analyzed, including those of fungi. Among these, residues 135 to 142 had been proposed to constitute the active site (Henrissat 1990). The presence of cysteine residues generally corresponded to an invariant position within the alignment (Fig. 2). The conserved cysteine residues form three disulfide bridges in hevamine (Jekel et al. 1991; Terwisscha van Scheltinga et al. 1993) and are most likely involved in conferring a (β/α) eight-barrel fold (Coulson 1994; Terwisscha van Scheltinga et al. 1994).

A nucleotide sequence matrix corresponding to the conserved domain was used for the neighbor-joining analysis of class III–related sequences (Fig. 3A). Essentially the same groupings were supported by the parsimony analysis (bold lines in Fig. 3A). Slightly longer branches were obtained for fungal and concanavalin B sequences when compared to class III sequences. This analysis also resulted in an obvious clustering of yeast, zygomycetes, and fungal sequences separately from the plant sequences: This clustering was supported by a bootstrap value of 100% in both neighbor-joining and parsimony analyses. The distantly related concanavalin B appeared sister group to the coherent class III cluster supported by 98–100% of the bootstraps in both analyses. Other coherent groupings were observed between class III sequences from *N. tabacum*, from *Vigna unguiculata* and *A. thaliana*, from *Cicer arietinum* and *Psophocarpus tetragonolobus*, and among the three *Cucumis sativus* sequences (Fig. 3A). These last three sequences are located on the same gene cluster (GenBank Acc. No. M84214) and probably result from gene duplication events.

Relationships Between Class I, II, and IV Chitinases

The results of phylogenetic analyses for 43 class I, II, and IV chitinase gene sequences are presented in Fig. 3B. Three main clusters were supported by the neighbor-joining and parsimony analyses. The first one contained eight class IV sequences while the other two comprised both class I (including I*) and class II sequences originating either from monocot (10 sequences) or from dicot species (25 sequences). Within the class IV cluster, the sequences from the monocot *Zea mays* formed a well-supported sister group to the dicot sequences. However, there was no well-supported division between sequences encoding basic isoforms (from *Z. mays*, *Beta vulgaris*-a, *Sambucus nigra*-a and *B. napus*) and acidic isoforms (from *Phaseolus vulgaris*, *S. nigra*-b and *B. vulgaris*-b). In the second main cluster comprising class I–II sequences from monocots—represented by four taxa from the Poaceae—a supported grouping contained all but one of the typical class I (i.e., CTE-encoding) sequences: This grouping suggested that these genes sharing similar structure were evolutionary related. On the other hand, two sequences from *H. vulgare* (class I* and class II), differing by the presence of a CRD coding region, appeared closely related according to a bootstrap value of 100% using both phylogenetic methods.

The third main cluster in the phylogenetic tree (Fig.

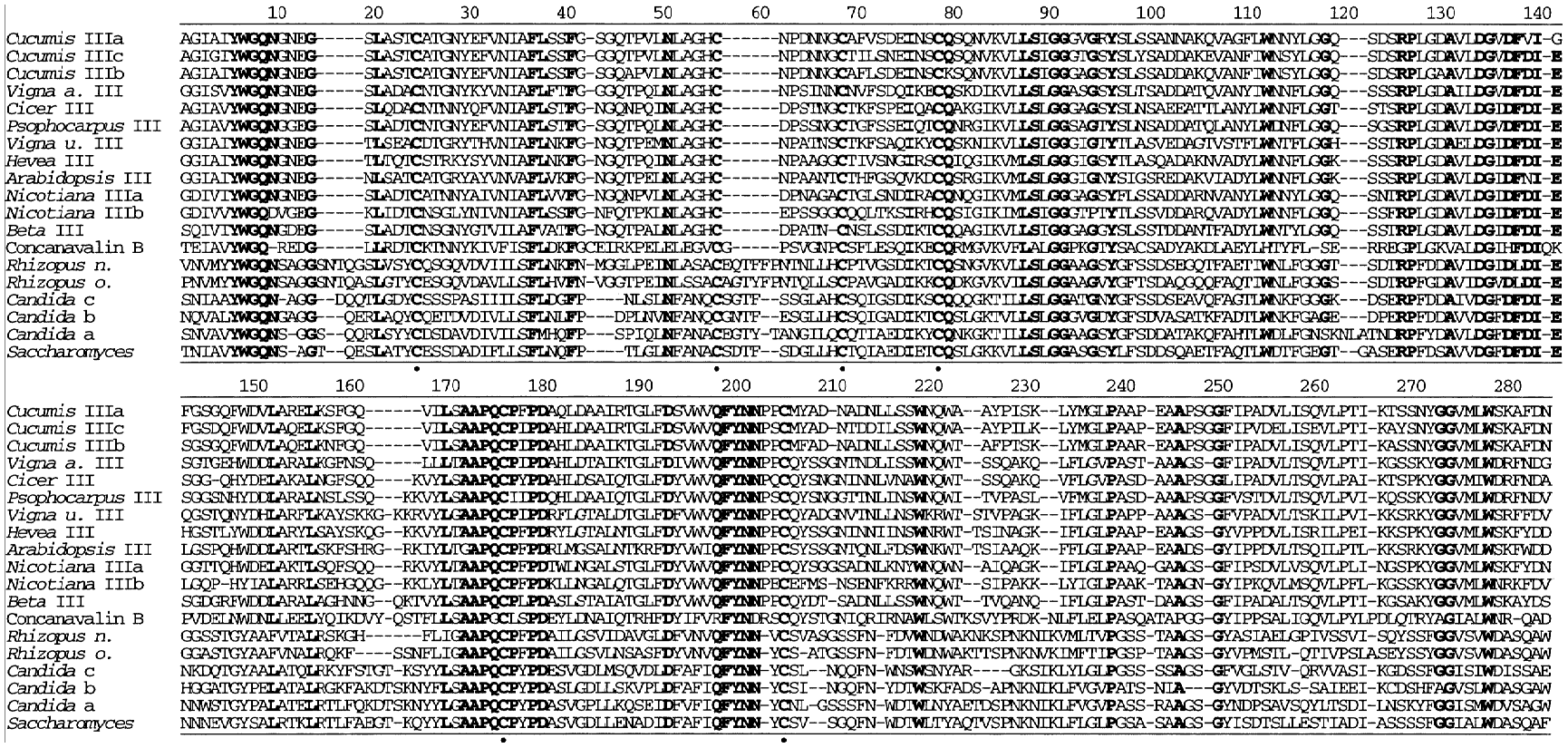

Fig. 2. Alignment of the partial amino acid sequences of class III plant chitinases, concanavalin B, and six related fungal chitinases. *Gaps* (-) were inserted to increase sequence identity. Residues in *boldface* are conserved in at least 17 out of the 19 sequences. Fungal sequences were obtained from GenBank database release 91 and are from *Saccharomyces cerevisiae* (M74069), *Candida albicans* (*a*, U15801; *b*, U15800; *c*, U36490), *Rhizopus niveus* (D10154), and *Rhizopus oligosporus* (D10157). The sequence of concanavalin is from *Canavalia ensiformis* (M83426), while sources for plant chitinases are listed in Table 1, except for *Hevea brasiliensis* (Jekel et al. 1991). The occurrence of cysteine residues at invariant positions of the alignment is indicated by *dots* below the alignment.

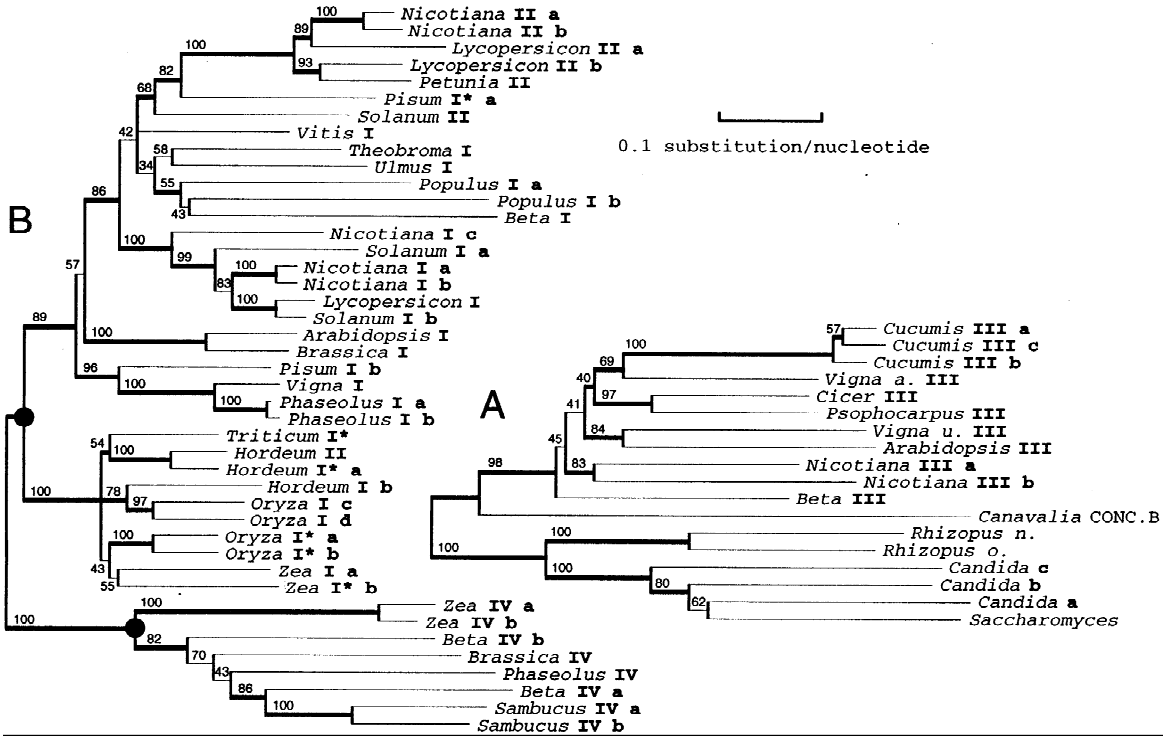

Fig. 3. Phylogeny reconstruction among 18 class III chitinases and related sequences **(A)** or among 43 class I, II, and IV chitinase sequences **(B)** inferred from neighbor-joining analyses of nucleotide substitution rates (Kimura two-parameter method). The nucleotide sequence data matrices contain 855 and 777 characters, respectively, whereas the overall standard errors on the two-parameter substitution rates are 6.9% and 7.6% for the first and the second data sets. The extent of the domains used for the phylogenetic analyses is specified in Materials and Methods. Phylogenic trees are drawn on the same scale for both data sets, while numbers indicate bootstrap estimates. Groupings supported by more than 50% of the bootstraps in parsimony analyses are shown by *bold horizontal lines.* Sequences are labeled by taxa, class designations, and in some cases by letters used for sequence differentiation (Table 1). *Filled circles* indicate the partition between sequences from monocots (Poaceae) and dicots. Sources of nucleotide sequences are listed in Table 1 and Fig. 2.

3B) contained most of the chitinase sequences analyzed, all obtained from dicot species. In agreement with established taxonomy, the following class I sequences appeared monophyletic with bootstrap values over 75%: the four Fabaceae sequences from *Pisum sativum, V. unguiculata,* and *P. vulgaris;* the two Brassicaceae sequences from *A. thaliana* and *B. napus;* the six Solanaceae sequences from *L. esculentum, N. tabacum,* and *S. tuberosum.* According to both phylogenetic analyses, the sequences from the Fabaceae and the Brassicaceae were supported sister groups to the other dicot sequences analyzed (Fig. 3B). Among the remaining group of sequences supported by a bootstrap value of 86% in the neighbor-joining analysis, most were typical class I sequences, the most frequent type recovered from angiosperm taxa. The class I sequences from *Populus trichocarpa* × *P. deltoides*-b and *B. vulgaris* showed considerably longer branches (Fig. 3B) and possibly high evolutionary rates. The last group of dicot sequences also comprised a cluster of five class II sequences from three different Solanaceae (*N. tabacum, L. esculentum,* and *P. hybrida*), which shared a common origin according to a bootstrap confidence level of 100% in both neighbor-joining and parsimony analyses. In a conserved region of the catalytic domain, these class II sequences also shared several residues that differed with respect to the corresponding residues in the other chitinases presented in Fig. 4. According to the phylogenetic reconstruction presented in Fig. 3B, the *N. tabacum, L. esculentum,* and *P. hybrida* class II sequences appear to be evolutionary related to a *S. tuberosum* class II gene and to a class I* sequence from a distant Fabaceae (*P. sativum*). The *S. tuberosum* class II gene shows similar structure but lacks the 14-codon deletion in the catalytic domain. In the case of the *P. sativum* sequence encoding a class I chitinase devoid of CTE, common ancestry closer than to any other sequence surveyed was supported by over 82% of the bootstraps in distance and parsimony analyses. A close relationship between class II and the *P. sativum* class I* sequences was also suggested by the amino acid alignment presented in Fig. 4 as well as by distance and parsimony analyses of amino acids (Hamel 1994; Araki and Torikata 1995). These observations were comparable to that on the class II sequence from *H. vulgare* (mentioned above) that also shared closer relationship with a class I* sequence than with any other class II sequence. Thus, even though all class II sequences possess a similar structure in lacking both types of auxiliary domains, they appear to be of polyphyletic origin when considered altogether.

Table 2. Substitution rates among groups of chitinase sequences analyzed in this study[a]

	d_K	P_S	P_A	P_S/P_A	tn/tv
C.I–II monocots	0.264 ± 0.050	0.382 ± 0.071	0.168 ± 0.038	2.43 ± 0.91	0.47 ± 0.17
C.I–II dicots	0.457 ± 0.115	0.709 ± 0.129	0.218 ± 0.064	3.72 ± 2.77	0.91 ± 0.34
C.IV	0.527 ± 0.115	0.752 ± 0.160	0.251 ± 0.045	2.97 ± 0.43	0.92 ± 0.10
C.I–II dicots vs monocots	0.555 ± 0.091	0.846 ± 0.103	0.237 ± 0.039	3.64 ± 0.64	0.89 ± 0.21
C.IV vs C.I–II monocots	0.755 ± 0.126	0.711 ± 0.182	0.383 ± 0.021	1.86 ± 0.50	0.79 ± 0.18
C.IV vs C.I–II dicots	0.790 ± 0.074	0.768 ± 0.074	0.389 ± 0.024	1.98 ± 0.25	0.83 ± 0.11

[a] d_K is the Kimura two-parameter distance, while P_S and P_A are proportions of synonymous and nonsynonymous substitutions per synonymous and nonsynonymous site, respectively. tn/tv is the ratio of transitions over transversions. Values are means ± SD of all pairwise comparisons between class I and II sequences from dicots (C.I–II dicots; 25 sequences) or monocots (C.I–II monocots; 10 sequences), or class IV sequences (C.IV; eight sequences)

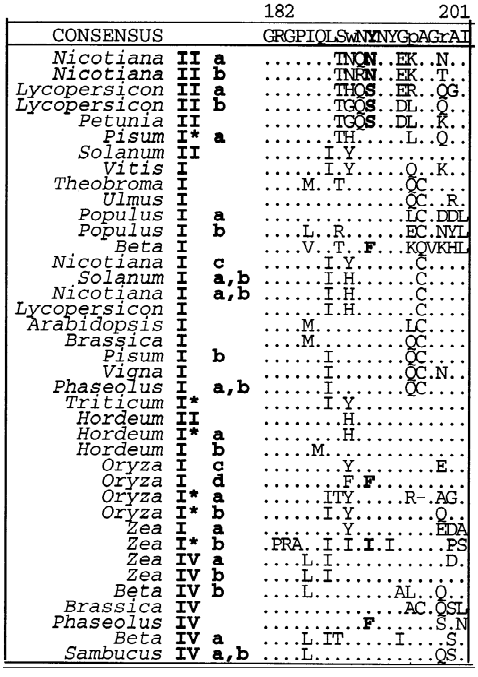

Fig. 4. Partial amino acid sequences for class I, II, and IV chitinases in an important region for the catalytic activity of *A. thaliana* class I chitinase (Verburg et al. 1993). Class labels are as in Fig. 3 and residues are *numbered* as in Hamel and Bellemare (1993). *Dots* denote identity with the consensus sequence. In the consensus sequence, residues conserved in 23 or more sequences are in *uppercase letters,* while the most frequent residues conserved in less than 23 sequences are in *lowercase.* A tyrosine at position 192 of the consensus sequence and the corresponding residues in the alignment are in *bold type.* These residues correspond to tyrosine-174 of Verburg et al. (1993) involved in substrate binding.

Evolutionary Divergence Among Class I, II, and IV Chitinases Gene Sequences

Various substitution rates were calculated within and between the three main clusters of class I, II, and IV sequences supported by the phylogenetic analyses (Fig. 3B). The fairly homogeneous class I and II monocot sequences differed from the other groups in showing considerably lower Kimura two-parameter distances and proportions of synonymous substitutions (Table 2). There was also a bias in favor of transversions (tv) in relation to transitions (tn), while no such bias was observed for the other comparisons. Similar values for the various parameters were obtained by comparing either within class I–II sequences from dicots, within class IV sequences, or between class I–II sequences from monocots and from dicots, suggesting a comparable level of evolutionary divergence. The number of synonymous substitutions was about 3.4-fold higher than nonsynonymous substitutions for these comparisons (Table 2). As could be inferred from identity levels and phylogeny reconstruction, a considerably larger evolutionary distance was observed between class IV and class I–II sequences than for all other comparisons. On the basis of substitution rates, class IV appeared equally distant from either the monocot or dicot clusters of class I–II sequences (Table 2). While proportions of synonymous substitutions remained fairly constant for all but one comparison in Table 2, greater proportions of nonsynonymous substitutions were obtained by comparing class IV with class I–II sequences. This suggests that the evolutionary divergence of class IV sequences resulted mainly from amino-acid-changing substitutions, hence modifying the primary structure and most likely the functional role of the chitinases encoded.

Discussion

Among the main group of chitinase sequences analyzed in this study, the largest, and therefore probably most ancient divergence is obtained by comparing class IV to class I and II sequences. According to many criteria including well-supported grouping in phylogenetic analyses, the distantly related group of class IV sequences, designated as class I-L by Araki and Torikata (1995), should rather deserve a separate class designation. Class I and class IV chitinases are found in both monocots and dicots without tandem clustering between class I and class IV sequences of the same taxonomic groups. The derivation of the class IV lineage from a common an-

cestral sequence would thus have occurred before the separation of monocots and dicots, estimated to have taken place around 200 million years ago (Wolfe et al. 1989). The remote divergence between class IV and class I–II sequences is reflected by a rooting of our phylogenetic tree (Fig. 3B), which is different from the tree presented by Araki and Torikata (1995). Our view differs correspondingly with the evolutionary scheme proposed by Araki and Torikata (1995), in which class IV derives from class I, which in turn would be derived from class II by insertion of the N-terminal CRD.

In contrast to class I and class IV sequences, this study provides no well-supported evidence for remote monophyletic origin for all class II sequences. When considered together, these sequences seem to constitute an artificial group, as at least one of the monocot sequences, from *H. vulgare*, has a separate origin from that of other class II sequences. The absence of auxiliary domains, a feature that usually forms the basis for defining class II, is indeed variable, even among closely related enzymes from the same plant species (Leah et al. 1991; Nishizawa et al. 1993; Beintema 1994). Sequences belonging to the main class II cluster, all obtained from Solanaceae species (*L. esculentum*, *N. tabacum*, and *P. hybrida*), possess the same conserved residues that have been shown to be important for catalysis (Glu 136[3] and Asn 193) and maintenance of the active site geometry (Thr 137, Gln 187) in class I and IV chitinases and other members of the lysozyme superfamily (Holm and Sander 1994; Hart et al. 1995). At the same time, these class II sequences share several particularities in a conserved region of the catalytic domain. Notably, Tyr 192 of class I and IV chitinases is substituted by a serine or an asparagine. The substitution of this tyrosine residue, thought to bind the substrate in the catalytic cleft (Hart et al. 1995), could reduce the catalytic activity in class II chitinases (Verburg et al. 1993). Another feature shared by these class II enzymes is the deletion of a 10-residue loop (165–181) located on the outside of the molecule, near the entrance of the catalytic cleft (Hart et al. 1995). In class I and IV enzymes, this loop contains a well-conserved tryptophan residue (Trp 172 in *B. napus*) thought to be important in substrate binding (Hart et al. 1995). The absence of this structure could thus modify interactions with the substrate, thereby modulating specificity. These observations are all consistent with results in *N. tabacum*, where class II enzymes show much lower specific activity than class I enzymes (Legrand et al. 1987), although this difference could also result from the absence of the chitin-binding CRD in class II enzymes (Raikhel et al. 1993; Graham and Sticklen 1994).

[3]Numbering of residues as found in BnCh25 (Hamel and Bellemare 1993)

Origin, Structure, and Function of Class III Chitinases

Class III proteins appear to be derived from an ancestral sequence different than that of class I, II, and IV plant chitinases, but similar to a type of chitinases found in yeast cells. The distribution of these related enzymes in both yeasts and angiosperms suggests that their putative ancestral sequence was present in a common parental lineage. According to this hypothesis, the origin of class III chitinases not only precedes the divergence between monocots and dicots, but also the division between fungi and plants.

In a study based on amino acid sequences that led to a general classification of all glycosyl hydrolases, Henrissat (1991) places class III chitinases in family 18 while class I and IV chitinases form family 19. Moreover, the secondary structure of class III enzymes is completely different from that of other plant chitinases hence they cannot be included in the lysozyme superfamily. They share a (β/α) eight-barrel topology also found in other glycosyl hydrolases and in some plant proteins that do not possess chitinase or lysozyme activities, such as concanavalin B (Hennig et al. 1995), the seed protein narbonin (Coulson 1994), and a broad bean nodulin (Perlick et al. 1996).

Class III chitinases also present distinct functional properties, since some of them possess a lysozyme activity (Jekel et al. 1991) not found in other classes of chitinases. Acidic and basic class III chitinases have been detected in the extracellular and intracellular cell fluids, respectively, of pathogen-infected plants (Métraux et al. 1989; Jekel et al. 1991). This suggests that they can be involved in defense against pathogens. However, the absence of sequence similarity with other pathogenesis-related chitinases and the fact that their related fungal counterparts are involved in cell separation during growth (Kuranda and Robbins 1991; McCreath et al. 1995) and in hyphal growth (Yanai et al. 1992) might indicate that plant class III enzymes have a distinct functional role. Instead of duplicating the functions of class I, II, and IV enzymes, they could be involved in morphogenesis during plant development and/or differentiation. A similar role in embryogenesis and differentiation has already been suggested in carrot for a 32-kDa acidic chitinase (De Jong et al. 1992), possibly a class III protein.

Evolutionary Trend Toward Deletion of Auxiliary Domains

Loss of the Cysteine-Rich Domain (CRD) Involved in Chitin Binding

The CRD has been previously shown to be essential for substrate binding, but not for catalytic or antifungal activity in a class I chitinase (Iseli et al. 1993). A modi-

fied CRD is also present in class IV chitinases and it seems most likely that a CRD was present in an ancestral sequence common to both classes. The three-dimensional structure of this domain has been maintained in various types of chitin-binding proteins such as hevein, wheat germ agglutinin, and plant chitinases: The eight cysteines thought to be involved in disulfide bridges (Beintema 1994) are very well conserved in most CRDs (Graham and Sticklen 1994).

It has been suggested that the CRD was introduced by some form of genetic transposition in the structure of a common ancestral gene that was probably similar to the present-day gene encoding hevein in *H. brasiliensis* (Shinshi et al. 1990; Araki and Torikata 1995). A similar transposition event might have led to the structure of two CRDs in tandem found in the amino-terminal region of a *Urtica dioica* protein sequence related to both lectins and chitinases (Lerner and Raikhel 1992).

Phylogenetic reconstructions presented in this study indicate that some of the chitinases devoid of CRD—namely, class II sequences—are derived from quite recent and independent deletion events that would have occurred in different taxonomic groups. For instance, the main cluster of class II sequences may be derived from an ancestral class I sequence devoid of CTE, similar to a present-day *P. sativum* sequence, by loss of the CRD. Similarly, the *H. vulgare* class II sequence analyzed in this study could result from another deletion event in a different class I or I* lineage. Data presented by Shinshi et al. (1990) strongly suggest that such excisions are favored by the presence of repeated DNA sequences at both ends of the CRD in class I sequences, the remains of which are still detectable in some class II sequences.

Loss of the Carboxy-Terminal Extension (CTE) Involved in Cellular Localization

The presence of a short CTE in most class I sequences has already been found to be necessary and sufficient for targeting chitinases to the plant vacuole (Neuhaus et al. 1991). In contrast, the absence of CTE is correlated with extracellular targeting for the class II and class IV chitinases whose cellular localization is known (de Tapia et al. 1987; Linthorst et al. 1990; Mikkelsen et al. 1992; Danhash et al. 1993). Extracellular, as opposed to intravacuolar chitinases, could be part of the foremost defense barrier against fungal pathogens. They could also perform signaling functions, releasing elicitors from either invading fungal hyphae (Mauch and Stahelin 1989; Graham and Sticklen 1994) or glycolipids present in the cell walls during development (De Jong et al. 1992).

In some of the plant species surveyed—namely, *B. napus*, *P. vulgaris*, and *Z. mays*—the presence of class IV chitinases fulfills the need for extracellular enzymes. On the other hand, class IV chitinases seem to have been replaced or complemented in various plant species by other types of extracellular chitinases. In the Solanaceae *N. tabacum* and *L. esculentum*, various class II but no class IV sequences have been isolated despite extensive surveys of their respective chitinase gene contents (Van Buuren et al. 1992; Danhash et al. 1993). Similarly, *O. sativa* and *Triticum aestivum*, two cereals cultivated for a long period, contain class I* sequences devoid of CTE, whereas no other type of extracellular chitinases has yet been identified in these species. Moreover, phylogenetic reconstructions argue in favor of repeated recent derivations of class I* chitinases, also present in *P. sativum* (Fabaceae), from ancestral class I genes found in different lineages of angiosperms.

There is considerable interest in obtaining more productive crop plants that also exhibit increased resistance to fungal pathogens. However, it is generally admitted that genetically uniform crop species often show increased susceptibility to various pathogens (Harlan 1975). The evolutionary schemes proposed in this study may contribute to a better understanding of the mechanisms involved in plant defense against fungal pathogens. Some of the structural changes between groups of evolutionarily related chitinases, such as the deletion of the CTE, have likely resulted in alteration of the host plant resistance. For instance, new types of extracellular chitinases could have arisen either by deletion or interruption of sequences encoding vacuolar targeting signals. At least some of these events are probably quite recent and seem to be restricted to discrete lineages of extensively cultivated plants. This raises the question of whether agriculture, through plant breeding and coevolution of crops with their associated pathogens, might contribute to the selection of individuals containing new types of extracellular chitinases.

Acknowledgments. We are grateful to Jean Bousquet (Centre de recherche en biologie forestière, Université Laval) and Lindsay Eltis (Département de biochimie, Université Laval) for critical review of the manuscript. This work was supported by National Sciences and Engineering Research Council of Canada operating grant A6923 and by the Conseil des recherches en pêche et en agro-alimentaire du Québec (CORPAQ) grant 3037 to G.B.

References

Araki T, Torikata T (1995) Structural classification of plant chitinases: two subclasses in class I and class II chitinases. Biosci Biotech Biochem 59:336–338

Beintema JJ (1994) Structural features of plant chitinases and chitin-binding proteins. FEBS Lett 350:159–163

Boller T (1988) Ethylene and the regulation of antifungal hydrolases in plants. In: Miflin BJ (ed) Surveys of plant molecular and cell biology, vol 5. Oxford University Press, Oxford, pp 145–174

Broglie KE, Chet I, Holliday M, Cressman R, Biddle P, Knowlton B, Mauvais CJ, Broglie R (1991) Transgenic plants with enhanced resistance to the fungal pathogen *Rhizoctonia solani*. Science 254: 1194–1197

Collinge DB, Kragh KM, Mikkelsen JD, Nielsen KK, Rasmussen U, Vad K (1993) Plant chitinases. Plant J 3:31–40

Coulson AF (1994) A proposed structure for ''family 18'' chitinases. A possible function for narbonin. FEBS Lett 354:41–44

Danhash N, Wagemakers CAM, van Kan JAL, de Wit PJGM (1993) Molecular characterization of four chitinase cDNAs obtained from *Cladosporium fulvum*-infected tomato. Plant Mol Biol 22:1017–1029

Davis JM, Clarke HRG, Bradshaw HD Jr, Gordon MP (1991) *Populus* chitinase genes: structure, organization, and similarity of translated sequences to herbaceous plant chitinases. Plant Mol Biol 17:631–639

De Jong AJ, Cordewener J, Lo Schiavo F, Terzi M, Vandekerckhove J, Van Kammen A, De Vries SC (1992) A carrot somatic embryo mutant is rescued by chitinase. Plant Cell 4:425–433

De Tapia M, Dietrich A, Burkard G (1987) *In vitro* synthesis and processing of a bean pathogenesis-related (PR4) protein. Eur J Biochem 166:554–563

Devereux J, Haeberli P, Smities O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Graham LS, Sticklen MB (1994) Plant chitinases. Can J Bot 72:1057–1083

Hamel F (1994) Isolement et caractérisation d'un gène de chitinase de *Brassica napus.* PhD Thesis, Université Laval, Québec Qc Canada

Hamel F, Bellemare G (1993) Nucleotide sequence of a *Brassica napus* endochitinase gene. Plant Physiol 101:1403

Harlan JR (1975) Les plantes cultivées et l'homme. Presses Universitaires de France, Paris

Hart PJ, Monzingo AF, Ready MP, Ernst SR, Robertus JD (1993) Crystal structure of an endochitinase from *Hordeum vulgare* L. seeds. J Mol Biol 229:189–193

Hart PJ, Pfluger HD, Monzingo AF, Hollis T, Robertus JD (1995) The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. J Mol Biol 229:189–193

Hennig M, Jansonius JN, Terwisscha van Scheltinga AC, Dijkstra BW, Schlesier B (1995) Crystal structure of concanavalin B at 1.65 Å resolution. An ''inactivated'' chitinase from seeds of *Canavalia ensiformis.* J Mol Biol 254:237–246

Henrissat B (1990) Weak sequence homologies among chitinases detected by clustering analysis. Protein Seq Data Anal 3:523–526

Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316

Holm L, Sander C (1994) Structural similarity of plant chitinase and lysozymes from animals and phages. FEBS Lett 340:129–132

Iseli B, Boller T, Neuhaus J-M (1993) The N-terminal cysteine-rich domain of tobacco class I chitinase is essential for chitin binding but not for catalytic or antifungal activity. Plant Physiol 103:221–226

Ishige F, Mori H, Yamazaki K, Imaseki H (1993) Cloning of a complementary DNA that encodes an acidic chitinase which is induced by ethylene and expression of the corresponding gene. Plant Cell Physiol 34:103–111

Jach C, Gornhardt B, Mundy J, Logemann J, Pinsdorf E, Leah R, Schell J, Maas C (1995) Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. Plant J 8:97–109

Jekel PA, Hartmann JBH, Beintema JJ (1991) The primary structure of hevamine, an enzyme with lysozyme/chitinase activity from *Hevea brasiliensis* latex. Eur J Biochem 200:123–130

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kragh KM, Jacobsen S, Mikkelsen JD, Nielsen KA (1993) Tissue specificity and induction of class I, II and III chitinases in barley (*Hordeum vulgare*). Physiol Plant 89:490–498

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA

Kuranda MJ, Robbins PW (1991) Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae.* J Biol Chem 266:19758–19767

Lawton K, Ward E, Payne G, Moyer M, Ryals J (1992) Acidic and basic class III chitinase mRNA accumulation in response to TMV infection of tobacco. Plant Mol Biol 19:735–743

Leah R, Tommerup H, Svendsen I, Mundy J (1991) Biochemical and molecular characterization of three barley seed proteins with antifungal properties. J Biol Chem 266:1564–1573

Legrand M, Kauffmann S, Geoffroy P, Fritig B (1987) Biological function of pathogenesis-related proteins: Four tobacco pathogenesis-related proteins are chitinases. Proc Natl Acad Sci USA 84: 6750–6754

Lerner DR, Raikhel NV (1992) The gene for stinging nettle lectin (*Urtica dioica* agglutinin) encodes both a lectin and a chitinase. J Biol Chem 267:11085–11091

Linthorst HJM, van Loon LC, van Rossum CMA, Mayer A, Bol JF, van Roekel JSC, Meulenhoff EJS, Cornelissen BJC (1990) Analysis of acidic and basic chitinases from tobacco and petunia and their constitutive expression in transgenic tobacco. Mol Plant Microbe Interact 3:252–258

McCreath KJ, Specht CA, Robbins PW (1995) Molecular cloning and characterization of chitinase genes from *Candida albicans.* Proc Natl Acad Sci USA 92:2544–2548

Margis-Pinheiro M, Metz-Boutigue MH, Awade A, de Tapia M, le Ret M, Burkard G (1991) Isolation of a complementary DNA encoding the bean PR4 chitinase: an acidic enzyme with an amino-terminus cysteine-rich domain. Plant Mol Biol 17:243–253

Mauch F, Stahelin A (1989) Functional implications of the subcellular localization of ethylene-induced chitinase and β-1,3-glucanase in bean leaves. Plant Cell 1:447–457

Meins F Jr, Neuhaus J-M, Sperisen C, Ryals J (1992) The primary structure of plant pathogenesis-related glucanohydrolases and their genes. In: Boller T, Meins F Jr (eds) Genes involved in plant defense. Springer Verlag, New York, pp 245–282

Melchers LS, Apotheker-de Groot M, van der Knaap JA, Ponstein AS, Sela-Buurlage MB, Bol JF, Cornelissen BJC, van den Elzen PJM, Linthorst HJM (1994) A new class of tobacco chitinases homologous to bacterial exo-chitinases displays antifungal activity. Plant J 5:469–480

Métraux JP, Burkhart W, Moyer M, Dincher S, Middlesteadt W, Williams S, Payne G, Carnes M, Ryals J (1989) Isolation of a complementary DNA encoding a chitinase with structural homology to a bifunctional lysozyme/chitinase. Proc Natl Acad Sci USA 86:896–900

Mikkelsen JD, Berglund L, Nielsen KK, Christiansen H, Bojsen K (1992) Structure of endochitinase genes from sugar beets. In: Brine CJ, Sandford PA, Zikakis JP (eds) Advances in chitin and chitosan. Elsevier, New York, pp 344–353

Monzingo AF, Marcotte EM, Hart PJ, Robertus JD (1996) Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eukaryotic families sharing a conserved core. Nat Struct Biol 3: 133–140

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Neuhaus J-M, Sticher L, Meins F Jr, Boller T (1991) A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. Proc Natl Acad Sci USA 88:10362–10366

Nielsen KK, Mikkelsen JD, Kragh KM, Bojsen K (1993) An acidic class III chitinase in sugar beet: induction by *Cercospora beticola,* characterization, and expression in transgenic tobacco plants. Mol Plant Microbe Interact 6:495–506

Nishizawa Y, Kishimoto N, Saito A, Hibi T (1993) Sequence variation, differential expression and chromosomal location of rice chitinase genes. Mol Gen Genet 214:1–10

Perlick AM, Fruhling M, Schroder G, Frosch SC, Puhler A (1996) The broad bean gene *VfNOD32* encodes a nodulin with sequence similarities to chitinases that is homologous to (alpha/beta)8-barrel-type seed proteins. Plant Physiol 110:147–154

Raikhel NV, Lee H-I, Broekaert WF (1993) Structure and function of chitin-binding proteins. Annu Rev Plant Physiol Plant Mol Biol 44:591–615

Rasmussen U, Bojsen K, Collinge DB (1992) Cloning and characterization of a pathogen-induced chitinase in *Brassica napus.* Plant Mol Biol 20:277–287

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Samac DA, Hironaka CM, Yallaly PE, Shah DM (1990) Isolation and characterization of the genes encoding basic and acidic chitinase in *Arabidopsis thaliana.* Plant Physiol 93:907–914

Schlumbaum A, Mauch F, Vogeli U, Boller T (1986) Plant chitinases are potent inhibitors of fungal growth. Nature 324:365–367

Shinshi H, Neuhaus J-M, Ryals J, Meins F Jr (1990) Structure of a tobacco endochitinase gene: evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain. Plant Mol Biol 14:357–368

Swofford DL (1993) PAUP, phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign, IL

Terwisscha van Scheltinga AC, Rozeboom HJ, Kalk KH, Dijkstra BW (1993) In: Muzzarelli RAA (ed) Chitin enzymology. Eur Chitin Soc, Ancona, pp 205–210

Terwisscha van Scheltinga AC, Kalk KH, Beintema JJ, Dijkstra BW (1994) Crystal structure of hevamine, a plant defense protein with chitinase and lysozyme activity, and its complex with an inhibitor. Structure 2:1181–1189

van Buuren M, Neuhaus J-M, Shinshi H, Ryals J, Meins F Jr (1992) The structure and regulation of homeologous tobacco endochitinase genes of *Nicotiana sylvestris* and *N. tomentosiformis* origin. Mol Gen Genet 232:460–469

Verburg JG, Rangwala SH, Samac DA, Luckow VA, Huynh QK (1993) Examination of the role of tyrosine-174 in the catalytic mechanism of the *Arabidopsis thaliana* chitinase: comparison of variant chitinases generated by site-directed mutagenesis and expressed in insect cells using baculovirus vectors. Arch Biochem Biophys 300:223–230

Vogelsang R, Barz W (1993) Cloning of a class III acidic chitinase from chickpea. Plant Physiol 103:297–298

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205

Yanai K, Takaya N, Kojima N, Horiuchi H, Ohta A, Takagi M (1992) Purification of two chitinases from *Rhizopus oligosporus* and isolation and sequencing of the encoding genes. J Bacteriol 174:7398–7406