

## Comparative Molecular Evolution of *Trichoderma* Chitinases in Response to Mycoparasitic Interactions

Katarina Ihrmark<sup>1</sup>, Nashwan Asmail<sup>1</sup>, Wimal Ubhayasekera<sup>2,4</sup>, Petter Melin<sup>3</sup>, Jan Stenlid<sup>1</sup> and Magnus Karlsson<sup>1</sup>

<sup>1</sup>Department of Forest Mycology and Pathology, Swedish University of Agricultural Sciences, Box 7026, S-75007, Uppsala, Sweden. <sup>2</sup>Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Center, Box 590, S-75124, Uppsala, Sweden. <sup>3</sup>Department of Microbiology, Swedish University of Agricultural Sciences, Box 7025, S-75007, Uppsala, Sweden. <sup>4</sup>MAX-lab, Lund University, Box 118, S-221 00 Lund, Sweden and Institute of Medicinal Chemistry, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen Ø, Denmark.  
Email: [magnus.karlsson@mykopat.slu.se](mailto:magnus.karlsson@mykopat.slu.se)

**Abstract:** Certain species of the fungal genus *Trichoderma* are potent mycoparasites and are used for biological control of fungal diseases on agricultural crops. In *Trichoderma*, whole-genome sequencing reveal between 20 and 36 different genes encoding chitinases, hydrolytic enzymes that are involved in the mycoparasitic attack. Sequences of *Trichoderma* chitinase genes *chi18-5*, *chi18-13*, *chi18-15* and *chi18-17*, which all exhibit specific expression during mycoparasitism-related conditions, were determined from up to 13 different taxa and studied with regard to their evolutionary patterns. Two of them, *chi18-13* and *chi18-17*, are members of the B1/B2 chitinase subgroup that have expanded significantly in paralog number in mycoparasitic *Hypocrea atroviridis* and *H. virens*. *Chi18-13* contains two codons that evolve under positive selection and seven groups of co-evolving sites. *Chi18-15* displays a unique codon-usage and contains five codons that evolve under positive selection and three groups of co-evolving sites. Regions of high amino acid variability are preferentially localized to substrate- or product side of the catalytic clefts. Differences in amino acid diversity/conservation patterns between different *Trichoderma* clades are observed. These observations show that *Trichoderma* chitinases *chi18-13* and *chi18-15* evolve in a manner consistent with rapid co-evolutionary interactions and identifies putative target regions involved in determining substrate-specificity.

**Keywords:** protein evolution, *Trichoderma*, mycoparasitism, chitinase

*Evolutionary Bioinformatics* 2010:6 1–26

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

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

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



## Introduction

Fungi are predominant pathogens on plants and infections have traditionally been controlled by chemical fungicides. Concerns about environmental impact of fungicides have made biological control an attractive option for managing plant diseases. Several mycoparasitic species of the anamorphic fungal genus *Trichoderma* are currently being used as biocontrol agents, e.g. *Trichoderma harzianum* (teleomorph *Hypocrea lixii*), *T. virens* (teleomorph *H. virens*), *T. atroviride* (teleomorph *H. atroviridis*) and *T. asperellum*.<sup>1</sup> For simplicity, we refer to *Trichoderma* and *Hypocrea* as *Trichoderma* in this study. *Trichoderma* spp. are frequently isolated from temperate and tropical soils, where they colonize woody and herbaceous material. Several mechanisms are proposed to be involved in the biocontrol ability of *Trichoderma* species, including direct mycoparasitic attack on plant pathogenic species, competition for plant exudates, nutrients or space, induced local and systemic response, and enhancement of plant growth.<sup>2</sup> The mycoparasitic attack often includes sensing and directed growth towards the antagonist,<sup>3</sup> followed by attachment and formation of appressoria. *Trichoderma* then secrete several cell wall degrading enzymes and mycotoxic peptaibol metabolites.<sup>4,5</sup> Chitin is an important constituent of fungal cell walls and chitinases have been shown to contribute to mycoparasitic attack.<sup>6</sup>

Fungal chitinases (EC3.2.1.14) exclusively belong to family 18 glycoside hydrolases and they are all predicted to possess a retaining mode of action.<sup>7</sup> Chitinases are involved in different biological functions such as cell wall remodelling during growth and development, degradation of chitin for nutritional needs and aggressive interactions with other fungi, insects and nematodes.<sup>8,9</sup> Whole genome sequencing of three different *Trichoderma* species, *H. jecorina*, *H. atroviridis* and *H. virens*, has revealed that a large number of chitinase genes are present in these species (20, 29 and 36 genes respectively).<sup>10,11</sup> Orthologs to the following chitinase genes have been previously cloned and studied from various *Trichoderma* species: *chi18-2*, *chi18-3*, *chi18-4*, *chi18-5*, *chi18-6*, *chi18-7*, *chi18-10*, *chi18-12*, *chi18-13*, *chi18-15* and *chi18-17*.<sup>10</sup> Based on previous work on phylogenetic relationships between chitinase catalytic domains, chitinases are divided into three main groups, A, B

and C. These groups are further subdivided into subgroups A2-A5, B1-B5 and C1-C2.<sup>7</sup> In certain *Trichoderma* chitinases, the catalytic domain is connected to substrate-binding domains which are not necessary for chitinolytic activity, but may enhance the efficiency of the enzymes.<sup>12,13</sup> Transcriptional patterns of *Trichoderma* chitinases show that some are expressed in response to mycoparasitic conditions, orthologs to *chi18-5*, *chi18-7*, *chi18-10*, *chi18-12*, *chi18-13*, *chi18-15* and *chi18-17*,<sup>10,14-18</sup> while others are constitutively expressed, *chi18-2*, *chi18-3* and *chi18-4*.<sup>10</sup>

Chitinase groups B and C are reported to display gene copy number expansions in some soil-borne ascomycetes, in addition to low levels of inter- and intraspecific amino acid conservation, which can be interpreted as a result of diversifying selection.<sup>7</sup> *Chi18-15* has been previously shown to be of actinobacterial origin and horizontally transferred to *Trichoderma*.<sup>12</sup> In addition, some plant defence chitinases from the genus *Arabidopsis* and the family *Poaceae* have evolved rapidly in response to a co-evolutionary arms race between plant host and fungal pathogen, resulting in a continuous selection for adaptive modifications.<sup>19,20</sup>

We hypothesize that *Trichoderma* chitinases, which have a function during the mycoparasitic interaction, have evolved adaptations to specific ecological contexts, such as cell wall composition of hosts, antagonistic microbial chitinase inhibitors and other environmental factors, of different *Trichoderma* species. We assume that specificity-determining residues have undergone mutations to compensate for the specificity needed; hence in paralogous or closely related orthologous sequences, specificity-determining residues may tend to display greater diversity than other positions. This concept was tested on four different *Trichoderma* chitinases that have been implicated in having a function during the mycoparasitic attack, *chi18-5*, *chi18-13*, *chi18-15* and *chi18-17*, by analyzing distribution of amino acid diversity, evolutionary rates and presence of co-evolving codons. In addition, we analyzed chitinase gene copy number expansions as the mycoparasitic lifestyle may have promoted the conservation of new paralogs.

We detected distinct differences in evolutionary trajectories that identify *chi18-13* and *chi18-15* as likely targets for adaptive evolution during mycoparasitic interactions. The results suggest that

fungal-fungal interactions can drive adaptive changes in enzymatic properties as a response to specific ecological contexts of different *Trichoderma* species.

## Materials and Methods

### Fungal material and media

Fifteen different strains of *Trichoderma* were used in the study (Table 1). Species assignment was based on analyses of species specific oligonucleotide barcodes located within the internal transcribed spacers 1 and 2 (ITS1 and ITS2) regions of the rRNA repeat, amplified by primers ITS1F and ITS4,<sup>21,22</sup> by using *TrichoKey* version 2.<sup>23</sup> Strains were maintained on malt extract agar at 25 °C.

### DNA extraction

For DNA extraction, conidia were transferred to liquid media (2% wv<sup>-1</sup> glucose, 2% wv<sup>-1</sup> yeast extract, 1% wv<sup>-1</sup> peptone) and incubated on a shaker for 16 to 54 hours. Mycelia were harvested, freeze-dried overnight, and crushed with a toothpick. An equal volume of DNA extraction buffer (0.2 M Tris (pH 8.5), 0.25 M NaCl, 0.5% SDS, 25 mM EDTA) was added and the mixes were incubated for 30 min at 70 °C followed by microcentrifugation at maximum speed for 5 minutes. Supernatants were RNase treated, extracted by phenol and chloroform followed by isopropanol precipitation.

**Table 1.** Fungal strains used in the current study.

Species	CBS number	Geographical origin
<i>H. citrina</i> <sup>a</sup>	593.76	Netherlands
<i>H. lixii</i> <sup>a</sup>	102174	Spain
<i>H. lixii</i> <sup>a</sup>	275.78	Colombia
<i>H. minutispora</i> <sup>a</sup>	341.93	Canada
<i>H. parapilulifera</i> <sup>a</sup>	112264	Australia
<i>H. pilulifera</i> <sup>a</sup>	224.84	Germany
<i>H. rufa</i> <sup>a</sup>	349.92	USA
<i>H. schweinitzii</i> <sup>a</sup>	258.85	USA
<i>H. virens</i> <sup>a</sup>	249.59	USA
<i>T. brevicompactum</i> <sup>a</sup>	109720	USA
<i>T. croceum</i> <sup>b</sup>	337.93	Canada
<i>T. ghanense</i> <sup>a</sup>	259.85	Canada
<i>T. longibrachiatum</i> <sup>a</sup>	182.69	Netherlands
<i>T. tomentosum</i> <sup>a</sup>	349.93	Canada
Unidentified <i>Trichoderma</i>	816.68	USA

<sup>a</sup>Based on ITS sequencing and *TrichoKey*.

<sup>b</sup>Based on morphological characters by Centraalbureau voor Schimmelcultures.

DNA pellets were dissolved and adjusted to 100 ng µl<sup>-1</sup> in 10 mM Tris (pH 8).

### Primer design and PCR

The coding regions of four chitinases, *chi18-5*, *chi18-13*, *chi18-15* and *chi18-17* were amplified from at least one strain per species, using primers listed in Supplemental Table S1. Sequences from the *H. jecorina*, *H. atroviridis* and *H. virens* genome projects (<http://www.jgi.doe.gov/>) were used for initial primer design; by aligning protein ID 80833 from *H. jecorina* and 111866 *H. virens* (*chi18-5*), 119859 *H. jecorina*, 45585 *H. atroviridis* and 25421 *H. virens* (*chi18-13*), 59791 *H. jecorina* and 89999 *H. virens* (*chi18-15*) and 110317 *H. jecorina* and 42107 *H. virens* (*chi18-17*) in the program BioEdit.<sup>24</sup> Primers were designed in conserved regions and evaluated with the program Primer3.<sup>25</sup> PCR was performed using 2720 and Veriti thermal cyclers (Applied Biosystems). The amplification was run with approximately 0.5 ng µl<sup>-1</sup> template DNA in a solution of 0.2 mM dNTP-mix, 0.03 U µl<sup>-1</sup> ThermoRed Taq DNA polymerase with buffer Y and enhancer P according to the manufacturer's instruction (Saveen and Werner) and 0.2–0.6 µM of each primer (the more degenerate the primer the higher the primer concentration) and 2.75 mM MgCl<sub>2</sub>. An initial denaturation step at 94 °C for 5 min was followed by 35 amplification cycles of denaturation at 94 °C for 30 s, annealing at 45–60 °C for 30 s and extension at 72 °C for 30–90 s. The thermal cycling was ended by a final extension step at 72 °C for 7 min. The PCR products were separated by gel electrophoresis on 1% agarose gels, purified with AMPure (Agencourt) and sequenced with a CEQ 8000 with the GenomeLab DTCS Quick Start Kit (Beckman Coulter). Sequence analyses and alignments were performed with the DNASTAR program package (DNASTar). Sequences were deposited in GenBank with accession numbers GU180607, GU186414-GU186439.

### Likelihood analysis of gene gain and loss

Using a species phylogeny and chitinase gene copy number in extant species as input data (Supplemental Fig. S1), the program CAFE (Computational Analysis of gene Family Evolution) was used to test for accelerated rates of chitinase family expansions or



contractions and identification of branches responsible for the non-random evolution.<sup>26,27</sup> Total number of chitinase genes for *H. atroviridis* was 29 (7 group A, 13 group B and 9 group C) and for *H. virens* 36 (8 group A, 13 group B and 15 group C).<sup>11</sup> In the species phylogeny (Fig. S1), *H. atroviridis* and *H. virens* were set as sister taxa, with a divergence time of 180 million years. They were in turn separated from *H. jecorina* by 10 million years. All additional species and divergence dates were as described previously.<sup>7</sup> The birth and death parameter ( $\lambda$ ) was estimated from the data and was 0.001 for all datasets. *P*-values were computed using 1000 re-samplings. Identification of the branch that was the most likely cause of deviations from a random model was determined by Viterbi and Likelihood ratio test procedures.<sup>26</sup> We considered *P*-values  $\leq 0.05$  or likelihood ratios above 50 to be significant for branch identification.

### Phylogenetic analysis

Translated gene products from the *H. atroviridis* and *H. virens* genome sequences were screened for the presence of chitinases using an iterative BLAST approach.<sup>7,28</sup> Amino acid sequences of chitinase catalytic domains were determined using InterProScan.<sup>29</sup> Sequences were aligned with Clustal W implemented in MEGA 4.0.2 using default settings and trimmed manually.<sup>30,31</sup> Phylogenetic analyses were performed using Neighbour-joining implemented in MEGA 4.0.2,<sup>31</sup> using either complete or pair wise deletion of gaps and missing data, and either a Poisson correction or the JTT substitution model.<sup>32</sup> Statistical support for phylogenetic grouping was assessed by 1000 bootstrap resamplings.

### Reverse conservation analysis (RCA)

From amino acid alignments of chitinases we identified regions of low conservation between closely related orthologs by applying RCA. Introns were removed from DNA sequences before translation. Amino acid sequences were aligned by Clustal X,<sup>33</sup> and RCA analyses were performed as described by Lee (2008).<sup>34</sup> In short, Rate4Site (Version 2.01) was used to calculate the degree of conservation (S score, high scores correspond to low degree of conservation) for each amino acid position using the empirical Bayesian method.<sup>35,36</sup> However, the S score varied considerably from residue to residue and was difficult to

analyze visually. Therefore, a sliding-window average ( $n = 7$ ) of normalized S scores (mean was 0 and standard deviation was 1) was plotted in Excel (Microsoft) (W mean score) and significant peaks were defined by intensity (I) values of 0.5 (i.e. 0.5 standard deviation), as recommended by Lee (2008).<sup>34</sup>

### Analysis of codon-usage

Differences in codon-usage between fungal and bacterial chitinases was analysed by performing multivariate (correspondence) analysis using the program CodonW,<sup>37</sup> accessed through the Mobylye web interphase (<http://mobylye.pasteur.fr/cgi-bin/portal.py?form=codonw>). In-frame, protein-encoding DNA sequences were used as input data. The universal genetic code was used, correspondence analysis was performed on codon usage frequencies, and all other options were set to default. Differences in codon-usage between genes were visualized by plotting the position of each gene on the resulting coa-axis 1 and 2 in Excel (Microsoft).

### Homology modelling of chitinases

Similar sequences were located in the protein entries of GenBank,<sup>38</sup> and aligned using Clustal W and hidden Markov models.<sup>30,39</sup> Family 18 chitinase catalytic domain structures were obtained from the Protein Data Bank (PDB),<sup>40</sup> then superimposed and compared with the programs LSQMAN and O.<sup>41,42</sup> Multiple sequence alignments were used to generate the best pair wise alignments, which were the basis for generating homology models of the catalytic modules of *H. jecorina* chitinases in the program SOD.<sup>41</sup> *Aspergillus fumigatus* chitinase (PDB entry 1W9P identity 60%),<sup>43</sup> was used as a template for modelling chi18-5, Chimerolectin from *Parkia platycephala* seeds (PDB entry 2GSJ.pdb identity 43%),<sup>44</sup> was used for chi18-13, *Streptomyces coelicolor* chitinase (PDB entry 3EBV.pdb (unpublished) identity 38%) was used for chi18-15 and *Saccharomyces cerevisiae* chitinase 1 (PDB entry 2UY2.pdb identity 47%),<sup>45</sup> was used for chi18-17. The models were adjusted in O, using rotamers that would improve packing in the interior of the protein and accounting for insertions and deletions in loop regions. The models are available upon request from the authors. The figure was prepared using MOLSCRIPT and Molray.<sup>46,47</sup>

## Codon-based analyses

Percent nucleotide identity between pair wise comparisons of chitinase genes were performed in MegAlign, implemented in the DNASTAR program package (DNASTAR).

The rate of non-synonymous (dN) and synonymous (dS) substitutions at each codon, and identification of sites evolving under positive or negative selection, was determined using the random effects maximum likelihood models (REL),<sup>48</sup> implemented in the HyPhy software package,<sup>49</sup> accessed through the Datamonkey webserver.<sup>50</sup> As recommended when using REL, the optimal nucleotide substitution model was estimated for each gene separately,<sup>48,51</sup> and included the following modifications to the general reversible nucleotide model:<sup>52–54</sup> *chi18-13*: C↔T: R<sub>CT</sub> and *chi18-15*: A↔C: R<sub>AC</sub>, A↔T: R<sub>AT</sub>, C↔G: R<sub>CG</sub>, C↔T: R<sub>CT</sub>, G↔T: R<sub>GT</sub>. A Bayes factor value ≥50 (default) was used as an indication of strong positive selection at a site, while values between 10 and 49 were considered to indicate weak support of positive selection.<sup>48</sup>

Identification of co-evolving sites was done using the Spidermonkey/BGM program implemented in the HyPhy software package,<sup>49,55</sup> accessed through the Datamonkey webserver.<sup>50</sup> The same nucleotide substitution models as were used for REL analysis were used. Global dN/dS values were estimated by the program, ambiguous characters were averaged, a two-parent, directed network was used and sites were selected based on non-synonymous branch counts (threshold ≥3). A posterior probability value ≥0.5 (default) was used as a definition of association between sites.

## Results

### Likelihood analysis of gene gain and loss

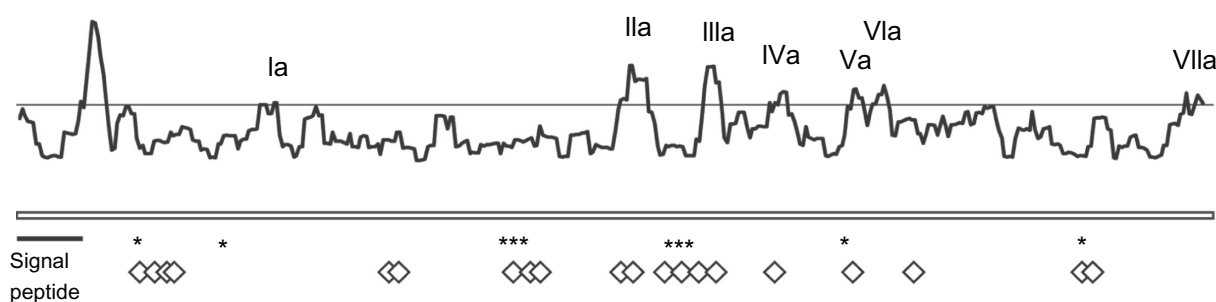
The size of the fungal chitinase gene family, including *H. jecorina*, *H. atroviridis* and *H. virens*, was tested for compatibility with a stochastic birth and death model using the program CAFE.<sup>26,27</sup> Previous results show that cluster A chitinases are closely related with cluster C and in order to assign expansions to chitinase subgroups, the data was analysed in three ways; cluster A + C chitinases separately, cluster B chitinase separately and all chitinases merged. The analyses showed that the fungal chitinase gene family,

analysing groups A + C and group B separately, as well as all chitinases merged, have evolved non-randomly ( $P \leq 0.015$ ) (Fig. S1). When analysing all chitinases merged, the branches for both *H. atroviridis* and *H. virens* were identified as contributing to the non-random pattern ( $P \leq 0.006$ ), compared with only the *H. virens* branch when groups A + C was analysed separately ( $P < 0.001$ ). The analysis of group B chitinases separately identified a non-random pattern for branches leading to *H. atroviridis*, *H. virens* and the ancestor to the *Trichoderma* clade as well as the ancestor to *H. atroviridis/H. virens* ( $P \leq 0.035$ ).

Analysis of gene phylogenies of chitinase subgroups identified subgroups C1 and C2 as the likely targets for the observed non-random gene copy number expansion in *H. virens* (Supplemental Fig. S2), as compared with other Sordariomycetes. *H. atroviridis* also contained a high number of C group chitinase genes, although the expansion was not statistically significant in the current analysis. Another observation was that while *H. virens* contained high numbers of both C1 and C2 chitinases, *H. atroviridis* contained mainly C1 chitinases and *H. jecorina* contained exclusively C2 members (Fig. S2). A more detailed analysis of group B chitinases revealed that the non-random expansion in the *Trichoderma* clade took place in the B1/B2 subgroup cluster (Fig. S2).

### Reverse conservation analysis of *chi18-5* amino acid variability

Amplification products and full-length sequences for *chi18-5* orthologs were successfully obtained from *H. schweinitzii*, *T. ghanense* and *T. longibrachiatum*. Additional sequences from *H. jecorina*, *H. atroviridis* and *H. virens* were retrieved from genome sequences and used for RCA analysis. A unique insert of 18 bp in *H. virens chi18-5* was excluded from the analysis. A phylogenetic analysis confirmed the orthologous status of the sequenced genes (Supplemental Figure S3). Amino acid diversity was distributed amongst eight regions with W mean scores above the 0.5 standard deviation threshold from the RCA analysis (Fig. 1). One of these regions was associated with the signal peptide cleavage site, while the other seven regions (Ia, IIa, IIIa, IVa, Va, VIa and VIIa) were visualized (Fig. 2A) using a homology model of *H. jecorina chi18-5*. Several of the twenty



**Figure 1.** Reverse conservation analysis of *chi18-5* orthologs. Amino acid diversity was estimated using Rate4Site, based on a Clustal X alignment of *chi18-5* *Trichoderma* orthologs, and plotted as W mean scores. The y-axis represents arbitrary units (not shown) while a horizontal line indicates a 0.5 standard deviation cutoff. The x-axis represents residue position, asterisks (\*) indicate positions of catalytic residues, diamonds (◇) indicate substrate-interacting residues. The positions of the signal peptide and regions with high amino acid diversity successfully visualised by homology modelling are indicated (Ia–VIIa).

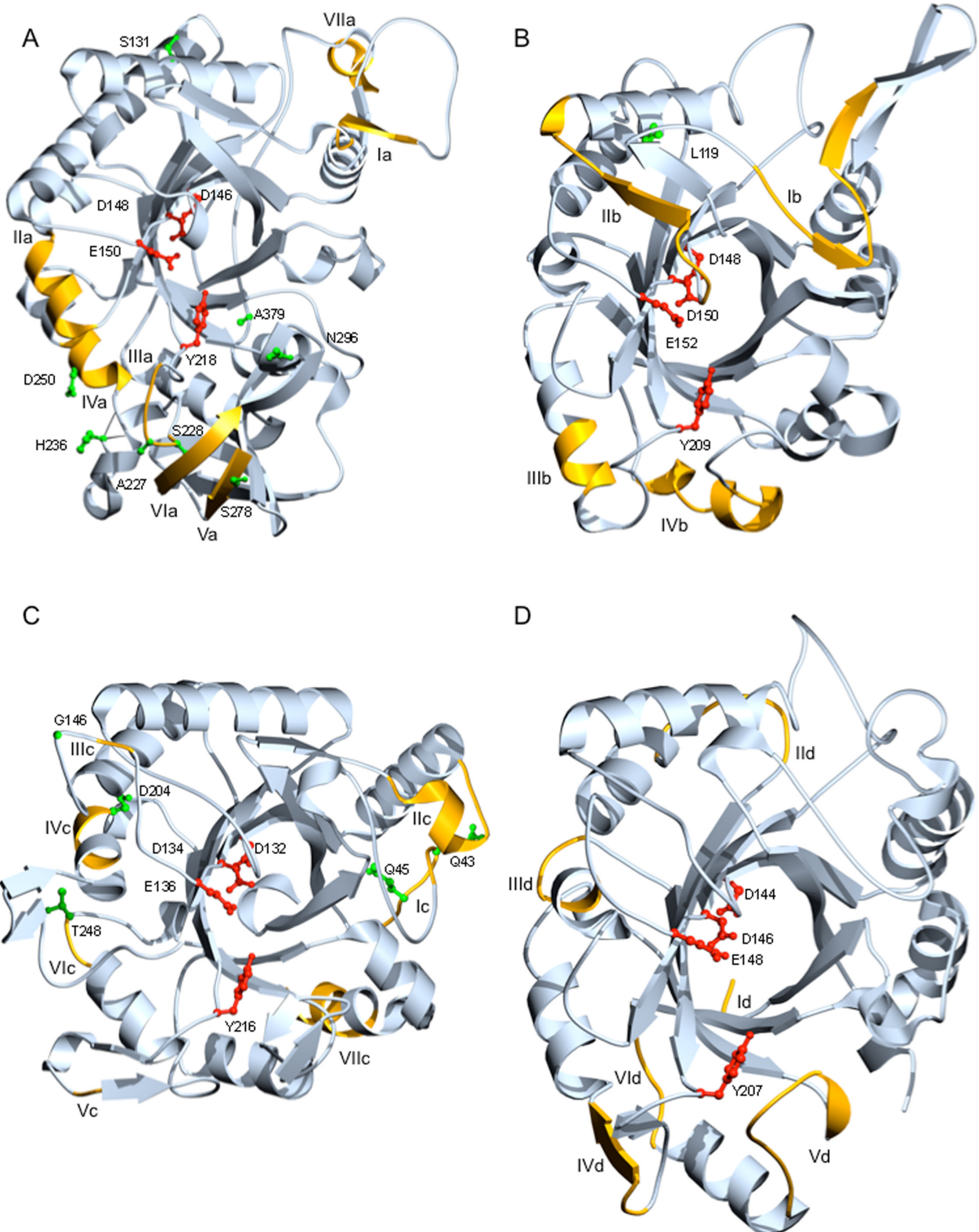
predicted residues (Supplemental Table S2) important for catalysis and substrate binding (cd06548 in Conserved Domain Database (CDD)<sup>56</sup>) were located in conserved regions with low W scores (Fig. 1). Three residues were located in regions with high W mean scores, one each in IIa, IIIa and IVa (Fig. 2A). These three regions, IIa, IIIa and IVa, were all surface-exposed and located on the product side of the enzyme. Region Ia was located in a loop that forms the entrance to the catalytic cleft, while region VIIa is on the surface of the enzyme but far from the catalytic site (Fig. 2A).

### Reverse conservation analysis of *chi18-13* amino acid variability

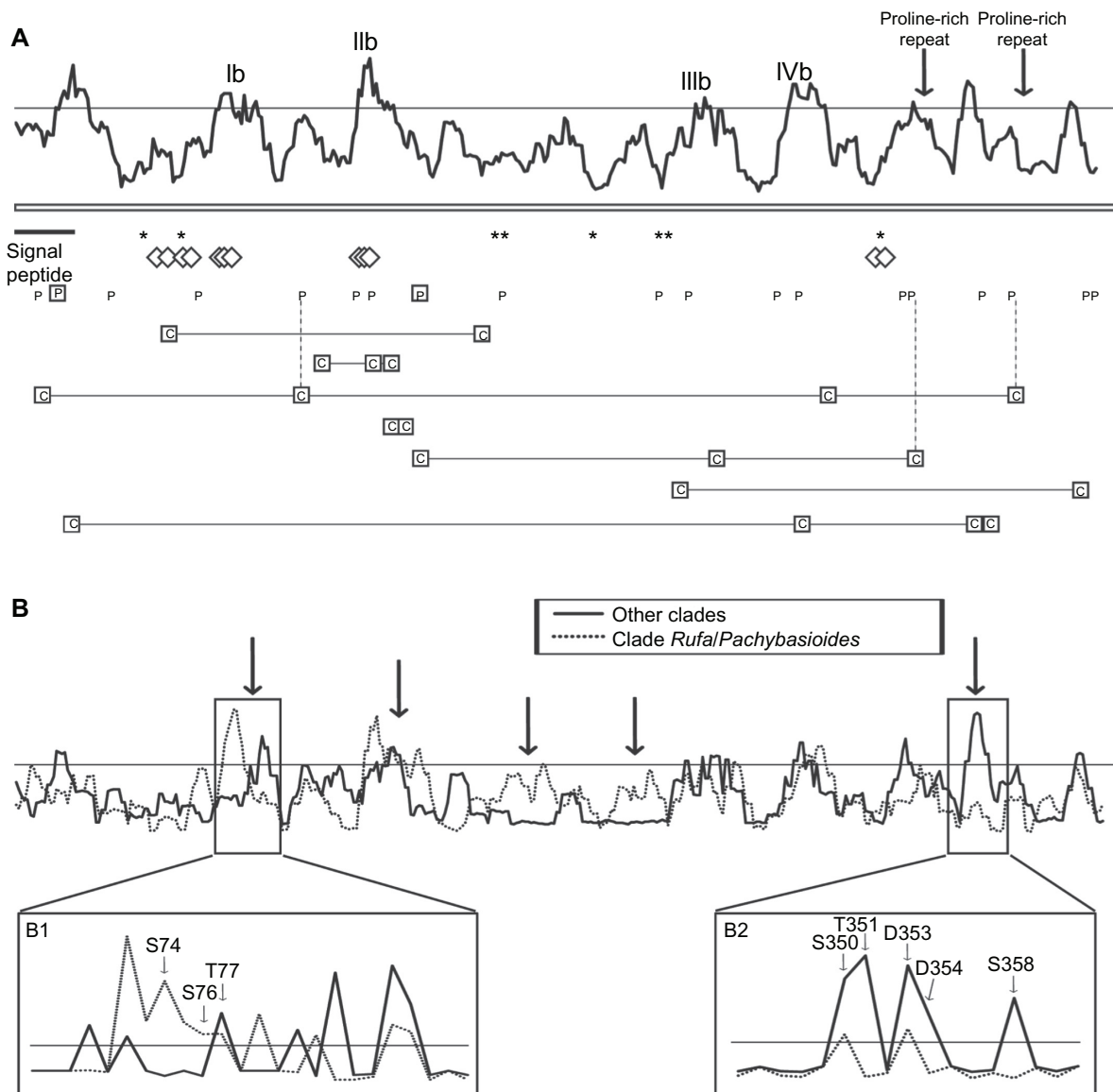
Amplification products and full-length sequences for the *H. schweinitzii* and *H. virens chi18-13* orthologs were obtained, as were partial sequences presumably lacking the eight C-terminal residues from an additional nine species (Fig. S3). The *H. virens chi18-13* gene was sequenced in the current work because the ortholog from the genome sequence (protein ID 25421) lacked 70 C-terminal residues compared to translated *chi18-13* orthologs from other *Trichoderma* species. Additional sequences from *H. jecorina* and *H. atroviridis* were retrieved from the genome sequences, together with two paralogous sequences, protein ID 79492 from *H. atroviridis* (originally cloned as Ech30)<sup>10,14</sup> and 58102 from *H. virens*. Two short proline-rich repeat regions in the C-terminal part (reference pos. 320–337 and 373–380 in *H. atroviridis* 45585) were removed from all species before the analysis, due to the highly variable number of repeats between species. The two paralogous sequences, 79492 and 58102,

were 70 C-terminal residues shorter than the orthologs. Phylogenetic analysis confirmed the orthologous and paralogous status of the selected sequences (Fig. S3). Amino acid diversity was distributed amongst eight regions with W mean scores above the 0.5 standard deviation threshold from the RCA analysis (Fig. 3A). Four regions (Ib, IIb, IIIb and IVb) were visualized by the homology model of *H. jecorina chi18-13* (Fig. 2B). One of the non-mapped high W score regions was associated with the signal peptide cleavage site (Fig. 3A). The eight predicted residues important for catalysis and substrate-binding by homology modelling (Table S2) were located in conserved regions with low W scores (Fig. 3A). Predicted substrate-binding residues (cd02877 in CDD) were associated with regions Ib and IIb of low amino acid conservation levels (Fig. 3A). More specifically, substrate-binding residues S<sub>74</sub>, S<sub>76</sub> and T<sub>77</sub> (reference *H. atroviridis*) were located in region Ib, which forms the entrance to the catalytic cleft (Fig. 2B). Substrate-binding residues G<sub>119</sub>, A<sub>120</sub> and V<sub>121</sub> (reference *H. atroviridis*) were located in region IIb, which forms a loop that protrudes into the catalytic centre of *chi18-13* (Fig. 2B). Regions IIIb and IVb were located on the surface but were not a part of the catalytic cleft (Fig. 2B).

Phylogenetic analysis of *chi18-13* (Fig. S3) revealed two separate groups of orthologs; one consisting of species from the taxonomic clades of *Rufa* and *Pashybasioides*,<sup>57</sup> the other consisting of species from several other *Trichoderma* clades. Although these groups were not always recovered in alternative phylogenetic analyses using other parameters such as the JTT substitution model in combination with complete deletion of missing data (data not shown), RCA analyses performed on



**Figure 2.** Homology models of *H. jecorina* chitinases. Homology models of the catalytic modules of *H. jecorina* chitinases (A) chi18-5, (B) chi18-13, (C) chi18-15 and (D) chi18-17 were generated using SOD and adjusted in O, based on hidden Markov models and Clustal W amino acid sequence alignments. Conserved catalytically important residues are indicated in red, amino acids under strong positive selection (Bayes factor  $\geq 50$ ) are indicated in green, variable regions from reverse conservation analysis (I scores  $\geq 0.5$ ) are indicated in orange and marked in Roman numerals from N- to C-termini. Residue numbering refers to catalytic module sequences used for modelling; 15–424 (chi18-5), 30–320 (chi18-13), 25–322 (chi18-15), 28–311 (chi18-17).



**Figure 3.** Reverse conservation analysis of chi18-13 orthologs and paralogs. **A)** Amino acid diversity was estimated using Rate4Site, based on a Clustal X alignment of chi18-13 *Trichoderma* orthologs and paralogs, and plotted as W mean scores. The y-axis represents arbitrary units (not shown) while a horizontal line indicates a 0.5 standard deviation cutoff. The x-axis represents residue position, asterisks (\*) indicate positions of catalytic residues, diamonds (◊) indicate substrate-interacting residues, boxed P indicate residues under strong (Bayes factor  $\geq 50$ ) positive selection, P indicate residues under weak (Bayes factor 10–49) positive selection, boxed C interconnected by horizontal lines indicate co-evolving residue groups and vertical dashed lines indicate identical residues. The position of the signal peptide, two proline-rich repeat units and regions with high amino acid diversity successfully visualised by homology modelling are indicated (Ib–IVb). **B)** Comparison of separate reverse conservation analyses on chi18-13 orthologs from *H. minutispora*, *H. parasilulifera*, *H. pilulifera* and *H. atroviridis* (dotted line) and *T. ghanense*, *H. jecorina*, *T. brevicompactum*, *H. citrina*, *H. schweinitzii*, *H. virescens*, *T. tomentosum*, *H. lixii* and strain CBS816.68 (solid line). Arrows indicate regions with different W mean score distribution, magnifications illustrate residue S score distribution of the selected region.

these two groups separately revealed five regions with amino acid conservation patterns indicative of differential adaptations between the two groups (Fig. 3B). A detailed analysis of the S scores for the individual amino acid positions revealed high S scores for substrate-interacting residues in region Ib in the *Rufa* and *Pachybasioides* clade (Fig. 3B). Low S scores were found for a region of unknown function situated between

two proline-rich-repeats in the C-terminal part of chi18-13 in the *Rufa* and *Pachybasioides* clade, compared with the other species (Fig. 3B).

### Reverse conservation analysis of chi18-15 amino acid variability

Amplification products and partial sequences for chi18-15 orthologs, presumably lacking 20–29 amino



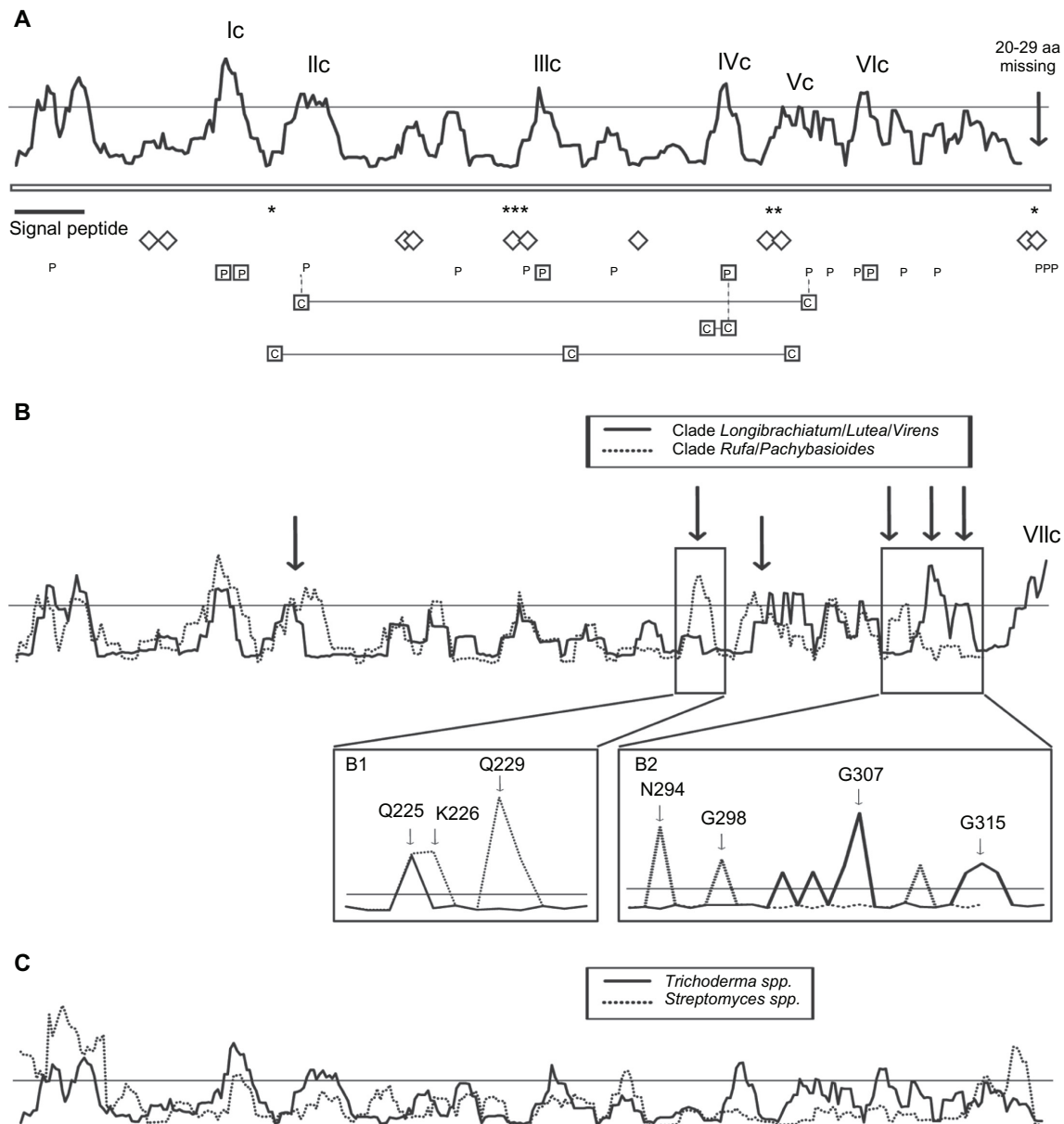
acid residues in the C-terminal part, were obtained from nine different *Trichoderma* species (Fig. S3). Additional full-length sequences from *H. jecorina*, *H. atroviridis* and *H. virens* were retrieved from the genome sequences. A phylogenetic analysis confirmed the orthologous status of the sequenced genes (Fig. S3). Amino acid diversity was distributed amongst eight regions with W mean scores above the defined threshold (Fig. 4A). Additional analyses of sequences from taxonomic subgroups (see below) identified an additional region of high amino acid diversity in the C-terminal end of chi18-15 (Fig. 4B). Two of these regions were associated with the secretion signal peptide while the other seven regions (Ic, IIc, IIIc, IVc, Vc, VIc and VIIc) were located in the catalytic module shown in Figure 2C prepared using the homology model of *H. jecorina* chi18-15. Although several of these regions were surface-exposed in the homology model, none of the parts contribute to the catalytic cleft. The seven residues predicted as important for catalysis in chi18-15 (cd02871 in CDD) were all located in conserved regions with low W scores, as were all predicted substrate-interacting residues (Fig. 4A).

Phylogenetic analysis of chi18-15 (Fig. S3) revealed two separate groups of orthologs; one consisting of species from the taxonomic clades of *Rufa* and *Pashybasioides*, the other consisting of species from clades *Longibrachiatum*, *Lutea* and *Virens*.<sup>57</sup> Again, some combinations of substitution models and handling of missing data in the phylogenetic analyses resulted in a more collapsed tree topology (data not shown), but the initial grouping was useful for more detailed analyses of amino acid variability. RCA analyses performed on these two groups separately revealed six regions with amino acid diversity patterns indicative of differential adaptations between the two groups (Fig. 4B). A detailed analysis of the S scores for the individual amino acid position revealed high conservation at position Q<sub>229</sub> (ref. *H. atroviridis*) in the *Longibrachiatum*, *Lutea* and *Virens* clades, where all species contained an aspartic acid residue while the position was occupied by either aspartic acid, glutamine or glutamic acid in the *Rufa* and *Pashybasioides* clade (Fig. 4B). Three other positions that displayed amino acid variability in either group were pos. N<sub>294</sub>, G<sub>298</sub> and G<sub>307</sub> (ref. *H. atroviridis*) in the C-terminal part of chi18-15 (Fig. 4B).

*H. jecorina* chi18-15 was previously shown to be of actinobacterial origin and horizontally transferred to *Trichoderma*,<sup>12</sup> most closely related to ChiJ from *S. coelicolor*. Therefore, a separate RCA analysis was performed on an alignment of six orthologs of ChiJ from *S. coelicolor*, *S. avermitilis*, *S. clavuligerus*, *S. ghanaensis*, *S. griseus* and *S. sp.* Mg1 (Supplemental Table S3). The result showed that the amino acid diversity in *Streptomyces* ChiJ was distributed at different position than among the *Trichoderma* chi18-15 orthologs (Fig. 4C).

### Reverse conservation analysis of chi18-17 amino acid variability

Amplification products and full-length sequences for *chi18-17* orthologs were successfully obtained from *H. schweinitzii*, *T. ghanense*, *T. tomentosum* and *H. lixii*, and additional sequences from *H. jecorina*, *H. atroviridis* and *H. virens* were retrieved from the genome sequences. Two introns were present in *H. jecorina*, *H. schweinitzii* and *T. ghanense*, compared with only one intron in *H. atroviridis*. No intron was present in *H. virens*, *T. tomentosum* or *H. lixii*, although in one isolate of *H. lixii* (CBS275.78) a unique insert was present that was interpreted as an intron and thus excluded from the analysis. A phylogenetic analysis confirmed the orthologous status of the sequenced genes (Fig. S3). Two short proline-rich repeat regions in the C-terminal part (pos. 324–331 and 350–355 in *H. atroviridis*) were excluded from the analysis, because the number of repeats was highly variable between species. Amino acid diversity was distributed amongst 13 regions with W mean scores above the defined threshold (Fig. 5). One of these regions was associated with the secretion signal peptide while two were situated in a C-terminal family 1 Carbohydrate-Binding Module (CBM1, cellulose and chitin binding), more specifically in the  $\beta$ 1 and  $\beta$ 2 antiparallel  $\beta$ -sheets (pos. 360–364 and 379–383 in *H. atroviridis*) (Fig. 5). Six regions were visualized using the homology model of *H. jecorina* chi18-17, (Id, IId, IIId, IVd, Vd and VIId) (Fig. 2D). Region Vd formed a part of the catalytic cleft while the other five were predicted to be surface-exposed but not directly associated with the catalytic cleft (Fig. 2D). The eight predicted residues important for catalysis and substrate-binding (cd02877 in CDD) were all located in conserved regions with low W scores (Fig. 5).

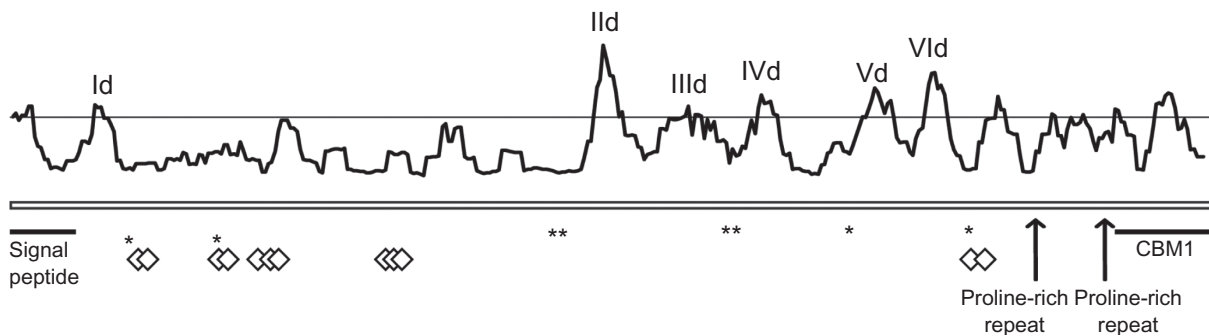


**Figure 4.** Reverse conservation analysis of *chi18-15* orthologs. **A)** Amino acid diversity was estimated using Rate4Site, based on a Clustal X alignment of *chi18-15* *Trichoderma* orthologs, and plotted as W mean scores. The y-axis represents arbitrary units (not shown) while a horizontal line indicates a 0.5 standard deviation cutoff. The x-axis represents residue position, asterisks (\*) indicate positions of catalytic residues, diamonds ( $\diamond$ ) indicate substrate-interacting residues, boxed P indicate residues under strong (Bayes factor  $\geq 50$ ) positive selection, P indicate residues under weak (Bayes factor 10–49) positive selection, boxed C interconnected by horizontal lines indicate co-evolving residue groups and vertical dashed lines indicate identical residues. The positions of the signal peptide, a C-terminal region not included in the overall analysis and regions with high amino acid diversity successfully visualised by homology modelling are indicated (Ic–VIc). **B)** Comparison of separate reverse conservation analyses on *chi18-15* orthologs from *H. minutispora*, *H. parasilulifera*, *H. pilulifera*, *H. atroviridis*, *H. rufa* and *T. croceum* (dotted line) and *T. ghanense*, *H. jecorina*, *T. brevicompactum*, *H. schweinitzii*, *H. virens* and *T. longibrachiatum* (solid line). Arrows indicate regions with different W mean score distribution, magnifications illustrate residue S score distribution of the selected region. **C)** Comparison of separate reverse conservation analyses on *chi18-15* orthologs from *Trichoderma* species and *ChiJ* orthologs from *Streptomyces* species.

## Analysis of codon-usage

Adaptation of codon-usage in *Trichoderma* chitinases and *Streptomyces* *ChiJ* orthologs was investigated by correspondence analysis of codon-usage using

CodonW.<sup>37</sup> Plotting coa-axis 1 and 2 for codon-usage for each gene identified three different clusters representing *chi18-15* orthologs, other *Trichoderma* chitinases and *Streptomyces* *ChiJ* orthologs (Fig. 6).



**Figure 5.** Reverse conservation analysis of *chi18-17* orthologs. Amino acid diversity was estimated using Rate4Site, based on a Clustal X alignment of *chi18-17* *Trichoderma* orthologs, and plotted as W mean scores. The y-axis represents arbitrary units (not shown) while a horizontal line indicates a 0.5 standard deviation cutoff. The x-axis represents residue position, asterisks (\*) indicate positions of catalytic residues, diamonds (◊) indicate substrate-interacting residues. The positions of the signal peptide, two proline-rich repeat units, a CBM1 substrate-binding region and regions with high amino acid diversity successfully visualised by homology modelling are indicated (Id-VIc).

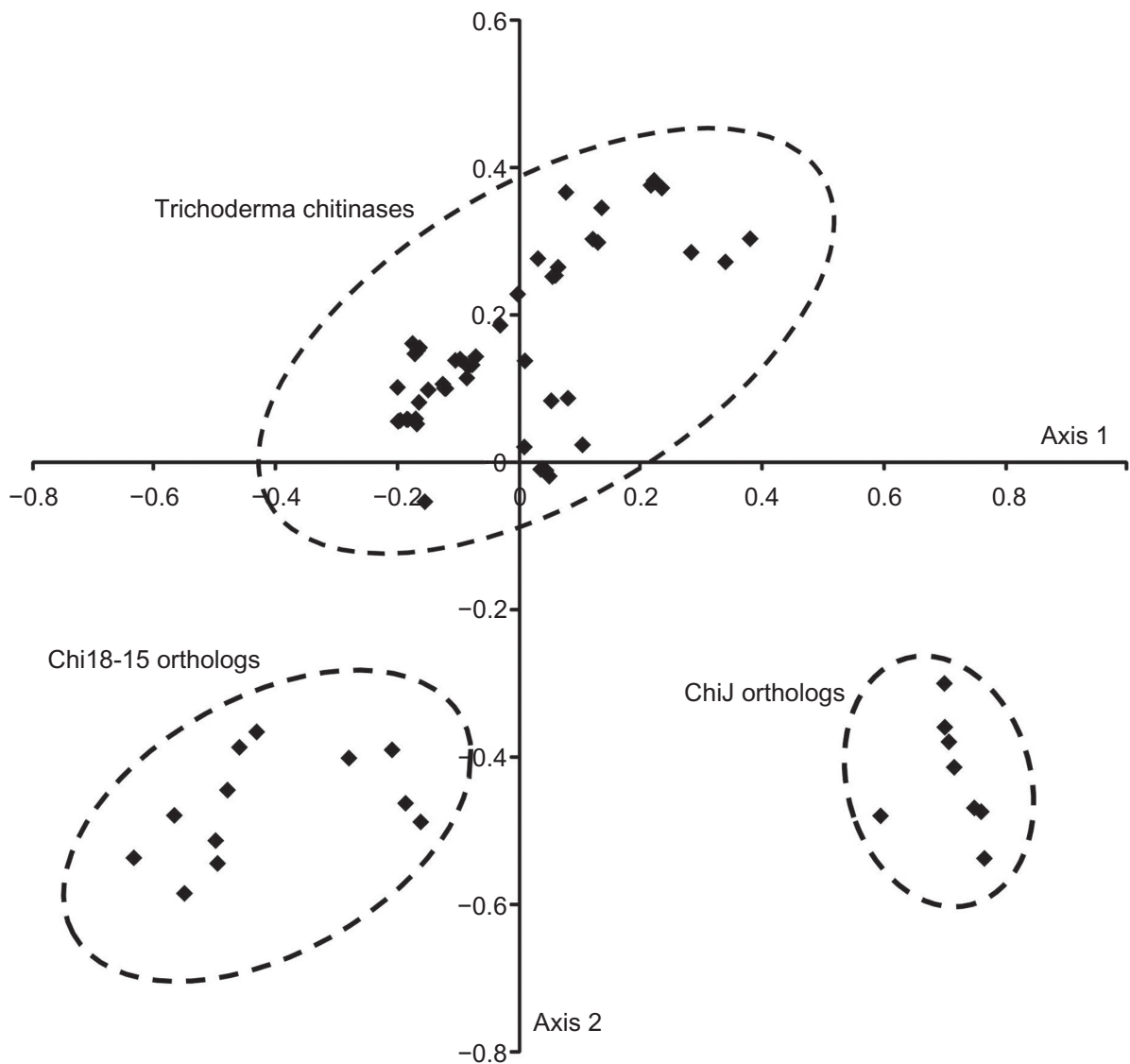
### Codon-based likelihood analyses

The mean pair wise nucleotide identity percentages among the four *Trichoderma* chitinases was;  $84.6 \pm 4.4$  (standard deviation) for *chi18-5*,  $88.6 \pm 4.0$  for *chi18-13*,  $80.3 \pm 6.1$  for *chi18-15* and  $83.1 \pm 6.3$  for *chi18-17*. Only *H. jecorina*, *H. atroviridis*, *H. virens*, *T. ghanense* and *H. schweinitzii* were included in this comparison, as data for only these species were available for all four genes.

In order to study the mechanisms behind the observed patterns of amino acid variability, we used REL analysis<sup>48</sup> to test for the presence of codons under different evolutionary constraints and to identify them. As recommended when using REL, ten species representatives were considered to be the minimum number of sequences for this analysis to provide reliable results; only *chi18-13* and *chi18-15* fulfilled this requirement. Two short proline-rich repeat regions in the C-terminal part of *chi18-13* were removed before REL analysis as the number of repeats was highly variable between species. REL fits both dN and dS substitution rates into three discrete distributions, yielding a total number of nine different rate classes of dN/dS. For *chi18-13*, one rate class was estimated to have dN/dS values above 1 (dN (0.49)/dS (0.21) = 2.33). A similar result was obtained for *chi18-15* where one rate class was estimated to have dN/dS values above 1 (dN (0.80)/dS (0.67) = 1.20). Using a cutoff of a Bayes factor  $\geq 50$ , two sites in *chi18-13* were identified as displaying signatures of positive selection (Table 2), 64 displayed signatures of

purifying selection, and 324 evolved neutrally. One of the positively selected sites was located in the signal peptide (pos. T<sub>16</sub>, ref. *H. atroviridis*). The other site was closely located to region IIb (pos. V<sub>137</sub>, ref. *H. atroviridis*), which was modelled to protrude into the catalytic centre of *chi18-13* (Fig. 2B). Furthermore, this second positively selected site coincided with one region with very different patterns of amino acid variability (W means) between *chi18-13* orthologs from the *Rufa/Pashybasioides* and other clades (Fig. 3A and 3B). Between a Bayes factor of 10 and 49, an additional 17 sites displayed weak signatures of positive selection (Fig. 3A). For *chi18-15*, using a cutoff of a Bayes factor  $\geq 50$ , five sites were identified as displaying signatures of positive selection (Table 2), 204 displayed signatures of purifying selection and 135 evolved neutrally. Between a Bayes factor of 10 and 49, an additional 13 sites displayed weak signatures of positive selection (Fig. 4A). All five positively selected sites were located in regions with high amino acid diversity, identified by RCA, pos. I<sub>68</sub> and K<sub>70</sub> (ref. *H. atroviridis*) in region Ic, pos. N<sub>171</sub> in IIIc, pos. Q<sub>229</sub> in IVc and pos. T<sub>273</sub> in VIc (Fig. 4A).

For comparative purposes, REL analysis was performed on partial sequences of two *Trichoderma* genes, the functions of which were assumed to be independent from mycoparasitic interactions, actin (*act*) and translation elongation factor 1 alpha (*tef*). These sequences were retrieved from GenBank (Table S3), and included 627 bp for *act* (pos. 2-628 in *H. virens*, acc. no. FJ442590) and 228 bp for *tef*



**Figure 6.** Codon-usage of *Trichoderma* chitinases and *Streptomyces ChiJ* orthologs. Correspondence analysis of codon-usage was performed on *Trichoderma chi18-5*, *chi18-13*, *chi18-15*, *chi18-17* orthologs and *Streptomyces ChiJ* orthologs, using the program CodonW accessed through the Mobyle web interphase. The resulting coa-axis 1 and 2 (in arbitrary units) for codon-usage for each gene was plotted in Excel. Dashed circles indicate groups of *chi18-15* orthologs, other *Trichoderma* chitinases and *ChiJ* orthologs.

(pos. 132–158, 258–320, 613–750 in *H. virens*, acc. no. EU280065). No sites displayed signs of positive selection (Bayes factor  $\geq 50$ ) in either gene.

### Analysis of co-evolving codons

Amino acid residues can interact structurally with each other to form and stabilize protein structures, or interact functionally through participation in the same protein function, such as substrate binding and processing. Therefore, co-evolution between sites in *chi18-13* and *chi18-15* were studied using evolutionary-network models implemented in Spidermonkey/BGM.<sup>55</sup> In *chi18-13* thirteen interacting pairs of

codons were identified (Table 3), forming seven groups of co-evolutionary sites (Fig. 3A). Three interacting groups included sites located in or close to region IIb, while three other groups included sites that were associated with regions IIIb or IVb. In three cases the interacting residues also showed weak signatures of positive selection (Fig. 3A). In *chi18-15*, four interacting pairs of codons were identified (Table 3), forming three groups of co-evolving sites (Fig. 4A). In one case, the interacting residue also displayed strong signatures of positive selection, and in two cases the interacting residues showed weak signatures of positive selection (Fig. 4A).

**Table 2.** Positively selected sites in chi18-13 and chi18-15.

Protein	Amino acid position <sup>a</sup>	Posterior probability <sup>b</sup>	Bayes factor <sup>b</sup>
Chi18-13	T <sub>16</sub>	0.88	52
	V <sub>137</sub>	0.97	213
Chi18-15	I <sub>68</sub>	0.69	62
	K <sub>70</sub>	0.80	111
	N <sub>171</sub>	0.81	119
	Q <sub>229</sub>	0.65	54
	T <sub>273</sub>	0.71	69

<sup>a</sup>Site position in reference to *H. atroviridis*.<sup>b</sup>Determined by Random Effects Likelihood method.

## Discussion

The complete genome sequence of three different *Trichoderma* species, *H. jecorina*, *H. atroviridis* and *H. virens* revealed the complexity of the chitinase enzyme system in these species illustrated by a total number of 20, 29 and 36 different chitinase genes respectively. The size of the chitinase gene family in the two mycoparasitic species *H. atroviridis* and *H. virens*, indicates that hydrolytic breakdown of the antagonists cell walls is important during the mycoparasitic interaction. However, chitinases are also

**Table 3.** Co-evolving sites in chi18-13 and chi18-15.

Protein	Amino acid position 1 <sup>a</sup>	Amino acid position 2 <sup>a</sup>	Posterior probability <sup>b</sup>	
Chi18-13	S <sub>51</sub>	T <sub>159</sub>	0.84	
	T <sub>97</sub>	A <sub>9</sub>	0.79	
	T <sub>97</sub>	N <sub>281</sub>	0.71	
	V <sub>121</sub>	Q <sub>125</sub>	0.63	
	S <sub>129</sub>	G <sub>130</sub>	0.86	
	S <sub>132</sub>	I <sub>312</sub>	0.64	
	S <sub>132</sub>	T <sub>241</sub>	0.72	
	G <sub>230</sub>	N <sub>369</sub>	0.90	
	Y <sub>270</sub>	S <sub>332</sub>	0.89	
	N <sub>281</sub>	K <sub>347</sub>	0.83	
	S <sub>332</sub>	L <sub>18</sub>	0.60	
	D <sub>335</sub>	L <sub>18</sub>	0.56	
	V <sub>121</sub>	K <sub>104</sub>	0.58	
	Chi18-15	S <sub>86</sub>	A <sub>178</sub>	0.59
		V <sub>90</sub>	A <sub>254</sub>	0.65
Q <sub>229</sub>		Q <sub>225</sub>	0.59	
S <sub>249</sub>		A <sub>178</sub>	0.56	

<sup>a</sup>Site position in reference to *H. atroviridis*.<sup>b</sup>For site 1 and site 2 to be conditionally dependent, determined by Spidermonkey/BGM.

involved in other functions such as morphological development, sporulation and autolysis.<sup>9</sup> Studying phenotypic effects in gene knock-out approaches is less likely to reflect the true biological function of a chitinase because of compensatory effects from paralogs, illustrating the need for complementary approaches.

Certain plant defence chitinases from the genus *Arabidopsis* and the family *Poaceae* evolved rapidly in response to a co-evolutionary arms race between plant host and fungal pathogen, resulting in a continuous selection for adaptive modifications.<sup>19,20</sup> Conceptually, the same situation may apply to microbe-microbe interactions; hence the combination of specific expression patterns during mycoparasitism and adaptive evolutionary changes may provide important information when assigning biological functions to *Trichoderma* chitinases.

## Chitinase gene family expansion

In the mycoparasitic species *H. atroviridis* and *H. virens* subgroups B1/B2 and C1/C2 have expanded significantly in paralog numbers. Stress-related genes often exhibit many expansions and contractions during fungal evolution;<sup>58</sup> hence the observed expansion suggest a role of at least some *Trichoderma* B1/B2 and C1/C2 subgroup members in aggressive fungal-fungal interactions. Gene duplications may relieve selective constraints on one gene copy which can evolve modified substrate specificities or enzyme properties more adapted towards specific cell wall composition in antagonistic species. Expansions of subgroups B1/B2 and C1/C2 are also observed in other soil-borne ascomycetes such as *Gibberella zeae*, *Uncinocarpus reesei* and *Emericella nidulans*, while gene copy number in these subgroups is reduced in the human pathogen *Coccidioides immitis*.<sup>7</sup>

## Evolution of chitinase chi18-13

*Chi18-13* is a member of the B1/B2 subgroup and displays the highest mean nucleotide conservation level among the studied chitinases. However, several codons are predicted to evolve under positive selection or form co-evolutionary site groups. Amino acid diversity is distributed amongst four successfully modelled regions, where Ib and Iib form parts of the catalytic cleft. This suggests that adaptations for substrate-specificity may be an important aspect



for *chi18-13* evolution. Three co-evolving groups of amino acid sites are associated with region IIb, supporting three substrate-interacting residues. It is possible that the observed site co-evolution is the result of modifications in the position of the substrate-interacting residues at optimal distances from the substrate in different *Trichoderma* species. In addition, three other interacting groups are associated with regions IIIb and IVb, and these co-evolving groups also include sites that are located in other parts of *chi18-13*, especially in the C-terminal part close to two proline-rich repeat units of variable lengths between species. These repeats can possibly function as linkers to provide flexibility in the secondary structure of *chi18-13*. This suggests that *chi18-13* processivity, in addition to substrate-specificity, has been under selection for modification during *chi18-13* evolution. Processivity may be influenced by distant parts of *chi18-13* and not merely by the loops that constitutes the catalytic cleft.

Expression data for the *H. atroviridis chi18-13* paralog *Ech30* (prot. ID 79492) show that the gene is induced by fungal cell wall material and during plate confrontation assays.<sup>10,14</sup> Enzyme activity of *H. atroviridis Ech30* shows that it is an endochitinase with preferential activity towards longer substrates, such as chitin fibre.<sup>14,59</sup> The low activity against short substrates suggests a shallow catalytic cleft for both *Ech30* and *chi18-13* which is in agreement with our modelling data. In summary, evolutionary data identify *chi18-13* as a candidate enzyme for mycoparasitic attack. *Chi18-13* is a member of a paralog cluster in mycoparasitic *H. atroviridis* and *H. virens* and displays signs of accelerated rates of evolution. Amino acid variability and co-evolution among sites of *chi18-13* are associated with regions that are not only predicted to influence substrate-specificity and processivity, but which also display differences in variability patterns between *Trichoderma* clades.

### Evolution of chitinase *chi18-15*

The optimal codon-usage of *chi18-15* is different from codon-usage in other *Trichoderma* chitinases and from the *Streptomyces ChiJ* gene. We previously demonstrated through phylogenetic analysis of the chitinolytic domain that *H. jecorina chi18-15* was introduced into the ancestor of *Trichoderma* through horizontal transfer from an actinobacterial origin.<sup>12</sup>

Two sites that evolved under positive selection are located in region Ic, which is located on the substrate entrance side of the protein, but not part of the catalytic cleft. In addition, two groups of co-evolutionary sites are located in both regions IIc and Vc which suggests concurrent structural adaptations of both substrate- and product sides of *chi18-15*. However, since neither region of high amino acid diversity (Ic to VIIc) is modelled to directly form part of the catalytic cleft, one interpretation is that they are the result of random mutations accumulating at regions with low selective constraints. The alternative explanation is that these regions are indeed important for functional properties of *chi18-15* and this explanation is supported for regions IVc and VIc. Both regions contain sites under accelerated evolutionary rates and display discrete differences between different *Trichoderma* clades or between *Streptomyces ChiJ* orthologs, which is not compatible with a completely stochastic process. In addition, region IVc contains two sites that co-evolved, again suggesting functional relevance for this region whether it be the maintenance of secondary structures, interaction with enhancer- or inhibitor proteins, or additional unknown functions.

Another region that displayed differences between *Trichoderma* clades contained site G<sub>298</sub> (ref. *H. atroviridis*) which is shown to determine activity inhibition by the chitinase inhibitor kinetin.<sup>45</sup> In *S. cerevisiae* CTS1, changing the alanine in this site to a bulkier serine residue eliminates inhibition by kinetin, whereas inhibition by allosamidin, acetazolamide and 8-chlorotheophylline remain unchanged.<sup>45</sup> It is possible that the observed differences between *Trichoderma* clades in this region represents an adaptation towards differences in antagonist inhibitor countermeasures. In comparison, the homologous position in *chi18-5*, *chi18-13* and *chi18-17* all contain a bulky methionine residue that may abolish inhibition.<sup>45</sup> This difference may reflect a recent introduction of *chi18-15* into *Trichoderma* and ongoing adaptations towards fungal preferences.

*Chi18-15* gene expression is induced by a variety of stimuli, including chitin and its monomers, nitrogen starvation, temperature and osmotic stress, and by interaction with *Rhizoctonia solani*.<sup>16</sup> *Chi18-15* has been reported to possess endochitinase activity with acidic optima and preferential activity towards high

molecular weight substrates.<sup>16,60</sup> As with *chi18-13*, evolutionary data identify *chi18-15* as a candidate enzyme for mycoparasitic attack. The gene has its ancestral origin as an actinobacterial chitinase, presumably an aggressive component in bacterial-fungal interactions. After being horizontally transferred into *Trichoderma*, it has been subjected to strong selective pressures to modify its functional properties according to the specific ecological contexts of different *Trichoderma* species.

### Evolution of chitinase *chi18-5*

In *chi18-5*, the observed amino acid diversity is preferentially situated in regions on the product side of the catalytic cleft that probably interact with substrate cleavage products as they leave the catalytic site. Therefore, these regions may be involved in determining the processivity of the enzyme. Regions IIa, IIIa and IVa are associated with residues that interact with the substrate at subsite +2, which has been confirmed by earlier studies from *T. harzianum chi18-5*.<sup>61</sup> The model also confirms the deep catalytic cleft of *chi18-5* which provides tight binding of the substrate.<sup>61</sup>

The *Chi18-5* orthologs are induced by chitin degradation products and during mycoparasitic interactions,<sup>10,15,18,62–65</sup> but gene knock-out experiments are inconclusive. Two reports show no reduction in the ability of *H. atroviridis* to overgrow other fungi in plate confrontation assays,<sup>66,67</sup> while one study showed a reduction of *H. virens* biocontrol ability.<sup>68</sup> Thus the function of *chi18-5* orthologs is suggested to be degradation of exogenous chitin for nutritional needs rather than a direct involvement in mycoparasitism.<sup>9</sup> This function does not contradict our evolutionary data; a conserved enzyme with a deep catalytic cleft that can bind chitin tightly. Slight modifications between different *Trichoderma* species may be associated with processivity but not with substrate specificity.

### Evolution of chitinase *chi18-17*

In *chi18-17*, the two regions of high amino acid diversity that are located in the CBM1 substrate-binding module suggest that there are discrete modifications of the binding properties of the *chi18-17* CBM1 domain in different *Trichoderma* species. A similar example is found in a plant defence chitinase where positively selected amino acid positions are located in a substrate-binding module.<sup>20</sup> Only

one variable region (Vd) is part of the catalytic cleft, which is wider and shallower than in *chi18-5*, while all other regions identified by RCA are predicted to be surface-exposed but not directly associated with the catalytic cleft. Interpretation of these regions is highly speculative; they may interact with other proteins in the environment, or alternatively they may represent regions of low selective constraints where a limited amount of amino acid variability is tolerated. *H. virens chi18-17* is reported to be induced by fungal cell wall material but in depth studies are missing.<sup>15</sup> The overall picture from evolutionary data depict a conserved protein with a shallow catalytic cleft, indicating endo-activity, without any obvious changes in known functional regions which suggests conserved enzymatic properties between *Trichoderma* orthologs.

### Concluding remarks

Certain plant defence chitinases from the genus *Arabidopsis* and the family *Poaceae* have evolved rapidly in response to a co-evolutionary arms race between plant host and fungal pathogen, resulting in a continuous selection for adaptive modifications.<sup>19,20</sup> Under these conditions, only a limited number of sites evolved under positive selection as severe structural constraints are present in chitinases to preserve catalytic function. This is observed also for fungal chitinases in the current study. In *chi18-13* and *chi18-15* only a few sites are identified to evolve under positive selection or to co-evolve with other sites. In addition, we detected a number of regions that display high amino acid diversity without any signs of accelerated evolution. Although one explanation may be low selective constraints in these regions, the localization to substrate- or product side of the catalytic cleft and differences in variability/conservation patterns between different *Trichoderma* clades, suggests that amino acid variation between species in at least some of these regions represents adaptive modifications.

The observed evolutionary differences between *chi18-5*, *chi18-13*, *chi18-15* and *chi18-17*, together with data on different domain-structures, different expression patterns and different enzymatic properties support the idea of functional differentiation between fungal chitinases. Therefore, correct functional assignment of individual genes and proteins are vital for a proper mechanistic understanding of biocontrol.



Using data on molecular evolution in a fungal-fungal interaction framework is one possible approach that can aid our understanding of mycoparasitism and structure/function relationships in enzymes.

## Acknowledgements

Genome sequence data were produced by the US Department of Energy (DOE) Joint Genome Institute (JGI) (<http://www.jgi.doe.gov/>). This work was supported by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS), the Carl Trygger Foundation, the Foundation for Strategic Environmental Research (PM), DANSCATT (WU), and The Royal Swedish Academy of Sciences. We thank Les Paul and Nick Rosenstock for linguistic revision.

## Disclosures

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors report no conflicts of interest.

## Abbreviations

CAFE, Computational analysis of gene family evolution; CBM, Carbohydrate-binding module; CDD, Conserved domain database; dN, Non-synonymous substitution rate; dS, Synonymous substitution rate; ITS, Internal transcribed spacer; MEGA, Molecular evolutionary genetics analysis; PDB, Protein data bank; RCA, Reverse conservation analysis; REL, Random effects maximum likelihood models.

## References

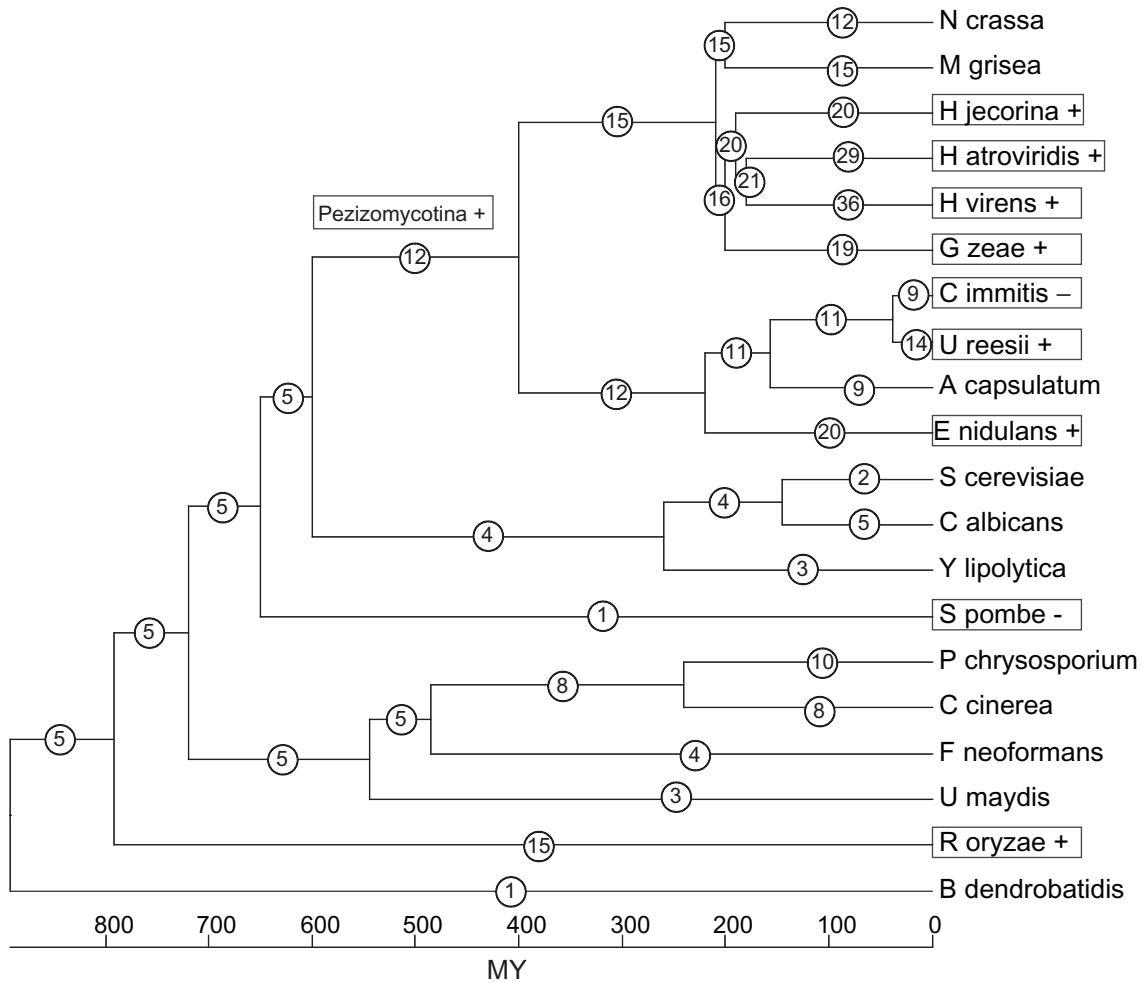
- Hjeljord L, Tronsmo A. In: Harman GE, Kubicek CP, editors. *Trichoderma and Gliocladium*. Taylor and Francis. 1998; p. 131–51.
- Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. *Trichoderma* species—Opportunistic, avirulent plant symbionts. *Nature Rev Microbiol*. 2004;2:43–56.
- Chet I, Harman GE, Baker R. *Trichoderma hamatum*—Its hyphal interactions with *Rhizoctonia solani* and *Pythium* spp. *Microb Ecol*. 1981; 7:29–38.
- Schirmböck M, Lorito M, Wang YL, et al. Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. *Appl Environ Microb*. 1994;60:4364–70.
- Howell CR. Mechanisms employed by *Trichoderma* species in the biological control of plant diseases: The history and evolution of current concepts. *Plant Dis*. 2003;87:4–10.
- Benitez T, Rincon AM, Limon MC, Codon AC. Biocontrol mechanisms of *Trichoderma* strains. *Int Microbiol*. 2004;7:249–60.
- Karlsson M, Stenlid J. Comparative evolutionary histories of the fungal chitinase gene family reveal non-random size expansions and contractions due to adaptive natural selection. *Evol Bioinform*. 2008;47–60.
- Duo-Chuan L. Review of fungal chitinases. *Mycopathologia*. 2006; 161:345–60.
- Seidl V. Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. *Fungal Biol Rev*. 2008;22:36–42.
- Seidl V, Huemer B, Seiboth B, Kubicek CP. A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of family 18 chitinases. *FEBS J*. 2005;272:5923–39.
- Seidl V, Gruber S, López-Mondéjar R, Zach S, Kubicek CP. The chitinolytic machinery of the mycoparasitic fungus *Trichoderma atroviride*. *Proc 9th Int Conf Eur Chitin Soc*. 2009. In press.
- Karlsson M, Stenlid J. Evolution of family 18 glycoside hydrolases: Diversity, domain structures and phylogenetic relationships. *J Mol Microbiol Biotechnol*. 2009;16:208–23.
- Limon MC, Margolles-Clark E, Benitez T, Penttilä M. Addition of substrate-binding domains increases substrate-binding capacity and specific activity of a chitinase from *Trichoderma harzianum*. *FEMS Microbiol Lett*. 2001;198:57–63.
- Klemsdal SS, Clarke JHL, Hoell IA, Eijsink VGH, Brurberg MB. Molecular cloning, characterization, and expression studies of a novel chitinase gene (*ech30*) from the mycoparasite *Trichoderma atroviride* strain P1. *FEMS Microbiol Lett*. 2006;256:282–9.
- Kim DJ, Baek JM, Uribe P, Kenerley CM, Cook DR. Cloning and characterization of multiple glycosyl hydrolase genes from *Trichoderma virens*. *Curr Genet*. 2002;40:374–84.
- Viterbo A, Montero M, Ramot O, et al. Expression regulation of the endochitinase *chit36* from *Trichoderma asperellum* (*T. harzianum* T-203). *Curr Genet*. 2002;42:114–22.
- de las Mercedes Dana M, Carmen Limón M, Mejías R, et al. Regulation of chitinase 33 (*chit33*) gene expression in *Trichoderma harzianum*. *Curr Genet*. 2001;38:335–42.
- Zeilinger S, Galhaup C, Payer K, et al. Chitinase gene expression during mycoparasitic interaction of *Trichoderma harzianum* with its host. *Fungal Genet Biol*. 1999;26:131–40.
- Bishop JG, Dean AM, Mitchell-Olds T. Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. *Proc Natl Acad Sci U S A*. 2000;97:5322–7.
- Tiffin P. Comparative evolutionary histories of chitinase genes in the genus *Zea* and family *Poaceae*. *Genetics*. 2004;167:1331–40.
- Gardes M, Bruns TD. Its primers with enhanced specificity for basidiomycetes—Application to the identification of mycorrhizae and rusts. *Mol Ecol*. 1993;2:113–8.
- White TJ, Bruns T, Lee S, Taylor J. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, Editors PCR protocols: A guide to methods and applications. San Diego: Academic; 1990;315–22.
- Druzhinina IS, Kopchinskiy AG, Komon M, et al. An oligonucleotide barcode for species identification in *Trichoderma* and *Hypocrea*. *Fungal Genet Biol*. 2005;42:813–28.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser*. 1999;41:95–8.
- Rozen S, Skaletsky H. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: Methods in molecular biology. Totowa: Humana Press; 2000;365–86.
- De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006;22:1269–71.
- Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res*. 2005;15:1153–60.
- Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
- Zdobnov EM, Apweiler R. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001;17:847–8.



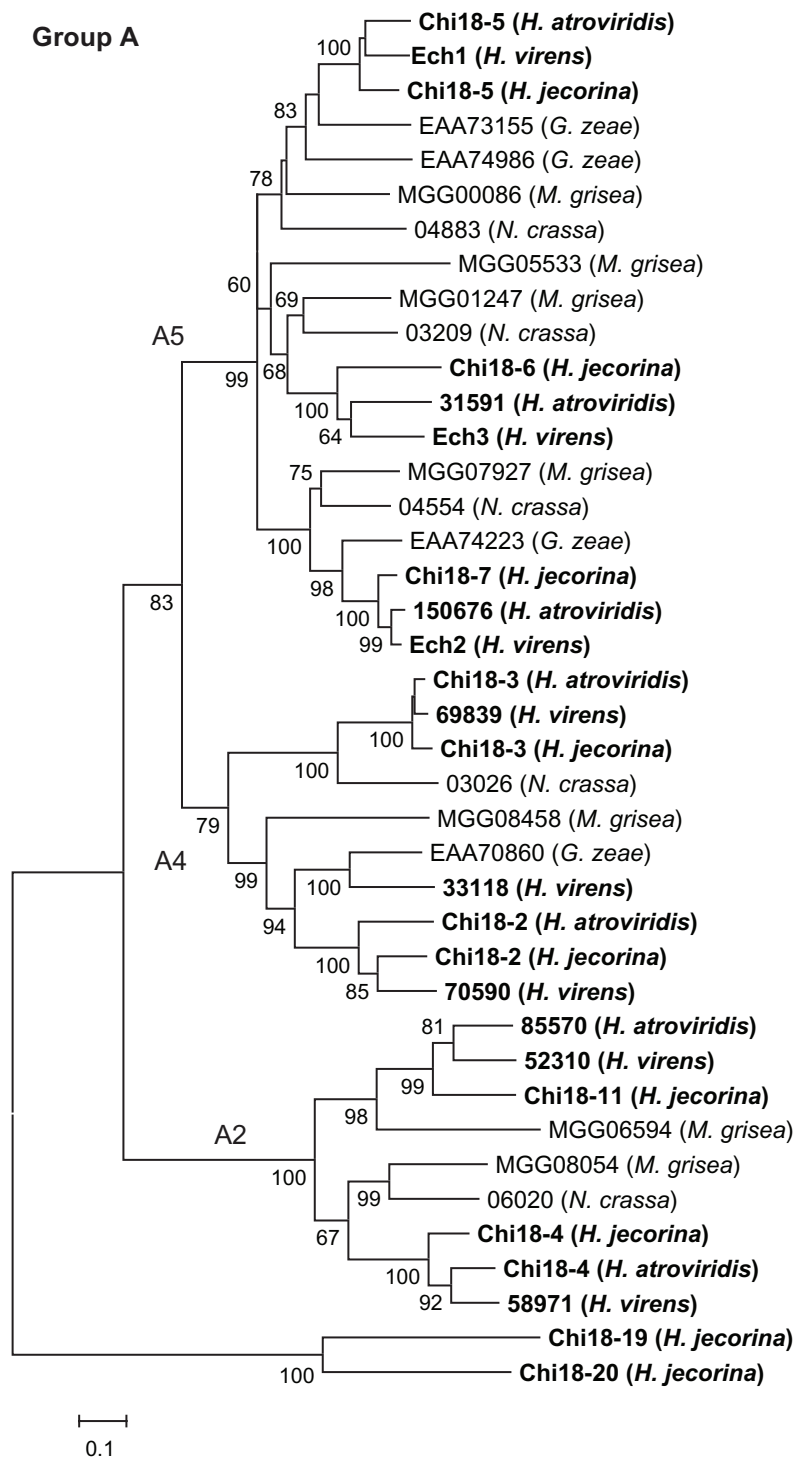


30. Thompson JD, Higgins DG, Gibson TJ. Clustal-W—Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
31. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007;24:1596–9.
32. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comp Appl Biosci.* 1992;8:275–82.
33. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
34. Lee T. Reverse conservation analysis reveals the specificity determining residues of cytochrome P450 family 2 (CYP 2). *Evol Bioinform.* 2008;4:7–16.
35. Mayrose I, Graur D, Ben-Tal N, Pupko T. Comparison of site-specific rate-inference methods for protein sequences: Empirical bayesian methods are superior. *Mol Biol Evol.* 2004;21:1781–91.
36. Pupko T, Bell R, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics.* 2002;18:S71–7.
37. Peden JF. Analysis of codon usage. Nottingham: University of Nottingham; 1999.
38. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2008;36:D25–30.
39. Karplus K, Katzman S, Shackelford G, et al. SAM-T04: what is new in protein-structure prediction for CASP6. *Proteins.* 2005;61:135–42.
40. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res.* 2000;28:235–42.
41. Kleywegt GJ, Zou JY, Kjeldgaard M, Jones TA. In: Rossmann MG, Arnold E, editors. International tables for crystallography, Vol. F. Crystallography of biological macromolecules. Dordrecht: Kluwer Academic; 2001:353–6,366–7.
42. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr A.* 1991;47:110–9.
43. Rao FV, Houston DR, Boot RG, et al. Specificity and affinity of natural product cyclopentapeptide inhibitors against *A. fumigatus*, human, and bacterial chitinases. *Chem Biol.* 2005;12:65–76.
44. Cavada BS, Moreno FBB, da Rocha BAM, et al. cDNA cloning and 1.75 angstrom crystal structure determination of PPL2, an endochitinase and N-acetylglucosamine-binding hemagglutinin from *Parkia platycephala* seeds. *FEBS J.* 2006;273:3962–74.
45. Hurtado-Guerrero R, van Aalten DMF. Structure of *Saccharomyces cerevisiae* chitinase 1 and screening-based discovery of potent inhibitors. *Chem Biol.* 2007;14:589–99.
46. Kraulis PJ. Molscript—a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr.* 1991;24:946–50.
47. Harris M, Jones TA. Molray—a web interface between O and the POV-Ray ray tracer. *Acta Crystallogr D.* 2001;57:1201–3.
48. Pond SLK, Frost SDW. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22:1208–22.
49. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;21:676–9.
50. Pond SLK, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21:2531–3.
51. Pond SLK, Frost SDW. A simple hierarchical approach to modeling distributions of substitution rates. *Mol Biol Evol.* 2005;22:223–34.
52. Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. *J Mol Evol.* 1984;20:86–93.
53. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci.* 1986;17:57–86.
54. Rodriguez F, Oliver JL, Marin A, Medina JR. The general stochastic-model of nucleotide substitution. *J Theor Biol.* 1990;142:485–501.
55. Poon A, Lewis F, Pond SLK, Frost S. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comp Biol.* 2007;3:e231.
56. Marchler-Bauer A, Anderson JB, Cherukuri PF, et al. CDD: a conserved domain database for protein classification. *Nucleic Acids Res.* 2005;33:D192–6.
57. Druzhinina I, Kubicek CP. Species concepts and biodiversity in *Trichoderma* and *Hypocrea*: from aggregate species to species clusters. *J Zhejiang Univ.* 2005;6B:100–12.
58. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature.* 2007;449:54–U36.
59. Hoell IA, Klemsdal SS, Vaaje-Kolstad G, Horn SJ, Eijsink VGH. Overexpression and characterization of a novel chitinase from *Trichoderma atroviride* strain P1. *Biochim Biophys Acta Prot Proteom.* 2005;1748:180–90.
60. Delacruz J, Hidalgalgallego A, Lora JM, et al. Isolation and characterization of 3 chitinases from *Trichoderma harzianum*. *Eur J Biochem.* 1992;206:859–67.
61. Boer H, Munck N, Natunen J, et al. Differential recognition of animal type beta 4-galactosylated and alpha 3-fucosylated chito-oligosaccharides by two family 18 chitinases from *Trichoderma harzianum*. *Glycobiology.* 2004;14:1303–13.
62. Haran S, Schickler H, Oppenheim A, Chet I. Differential expression of *Trichoderma harzianum* chitinases during mycoparasitism. *Phytopathology.* 1996;86:980–5.
63. Carsolio C, Gutierrez A, Jimenez B, Vanmontagu M, Herreraestrella A. Characterization of *Ech-42*, a *Trichoderma harzianum* endochitinase gene expressed during mycoparasitism. *Proc Natl Acad Sci U S A.* 1994;91:10903–7.
64. Garcia I, Lora JM, Delacruz J, et al. Cloning and characterization of a chitinase (*Chit42*) cDNA from the mycoparasitic fungus *Trichoderma harzianum*. *Curr Genet.* 1994;27:83–9.
65. Mach RL, Peterbauer CK, Payer K, et al. Expression of two major chitinase genes of *Trichoderma atroviride* (*T. harzianum* P1) is triggered by different regulatory signals. *Appl Environ Microbiol.* 1999;65:1858–63.
66. Carsolio C, Benhamou N, Haran S, et al. Role of the *Trichoderma harzianum* endochitinase gene, *ech42*, in mycoparasitism. *Appl Environ Microbiol.* 1999;65:929–35.
67. Woo SL, Donzelli B, Scala F, et al. Disruption of the *ech42* (endochitinase-encoding) gene affects biocontrol activity in *Trichoderma harzianum* P1. *Mol Plant Microbe In.* 1999;12:419–29.
68. Baek JM, Howell CR, Kenerley CM. The role of an extracellular chitinase from *Trichoderma virens* Gv29-8 in the biocontrol of *Rhizoctonia solani*. *Curr Genet.* 1999;35:41–50.

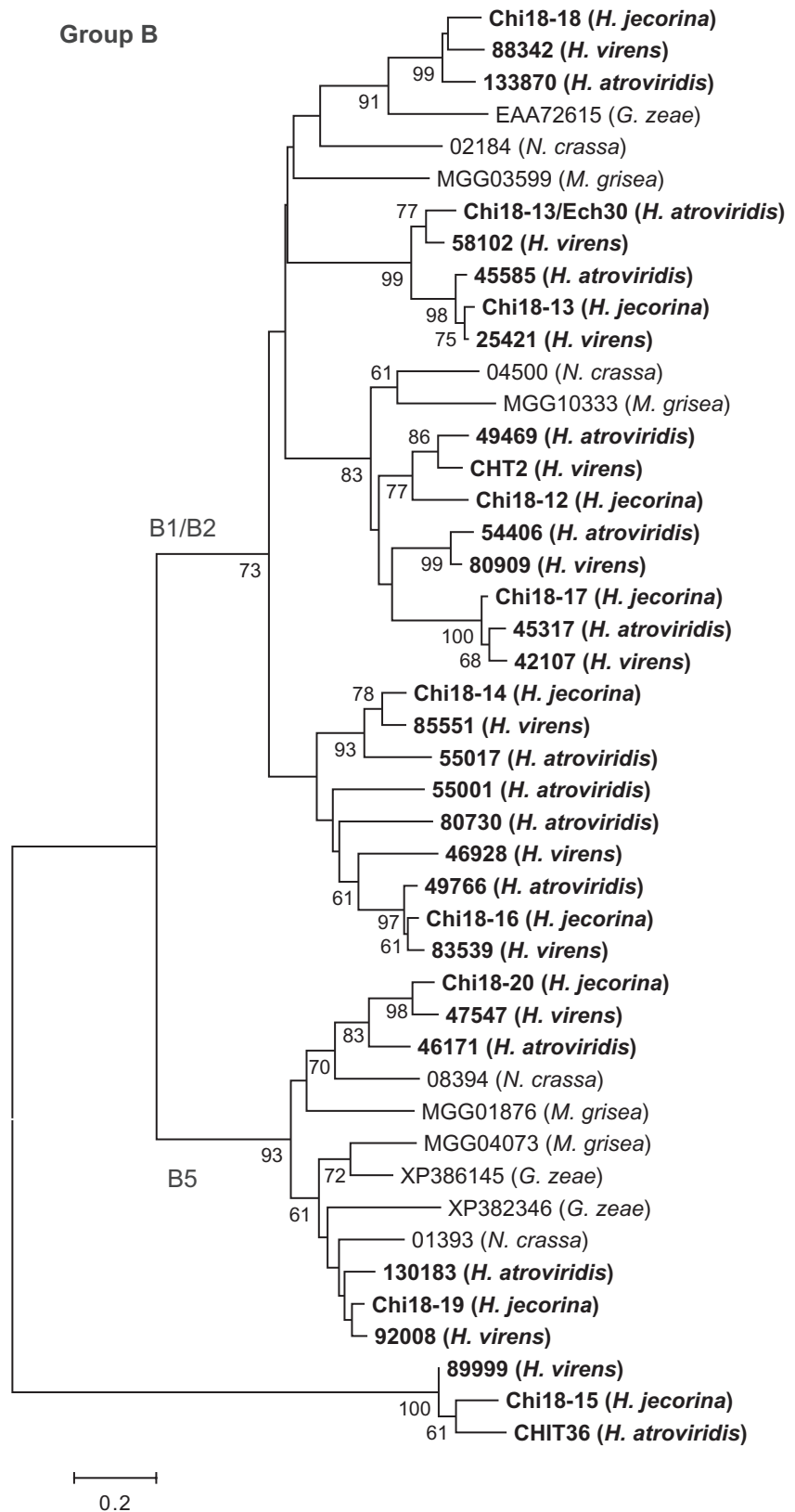
## Supplemental Materials



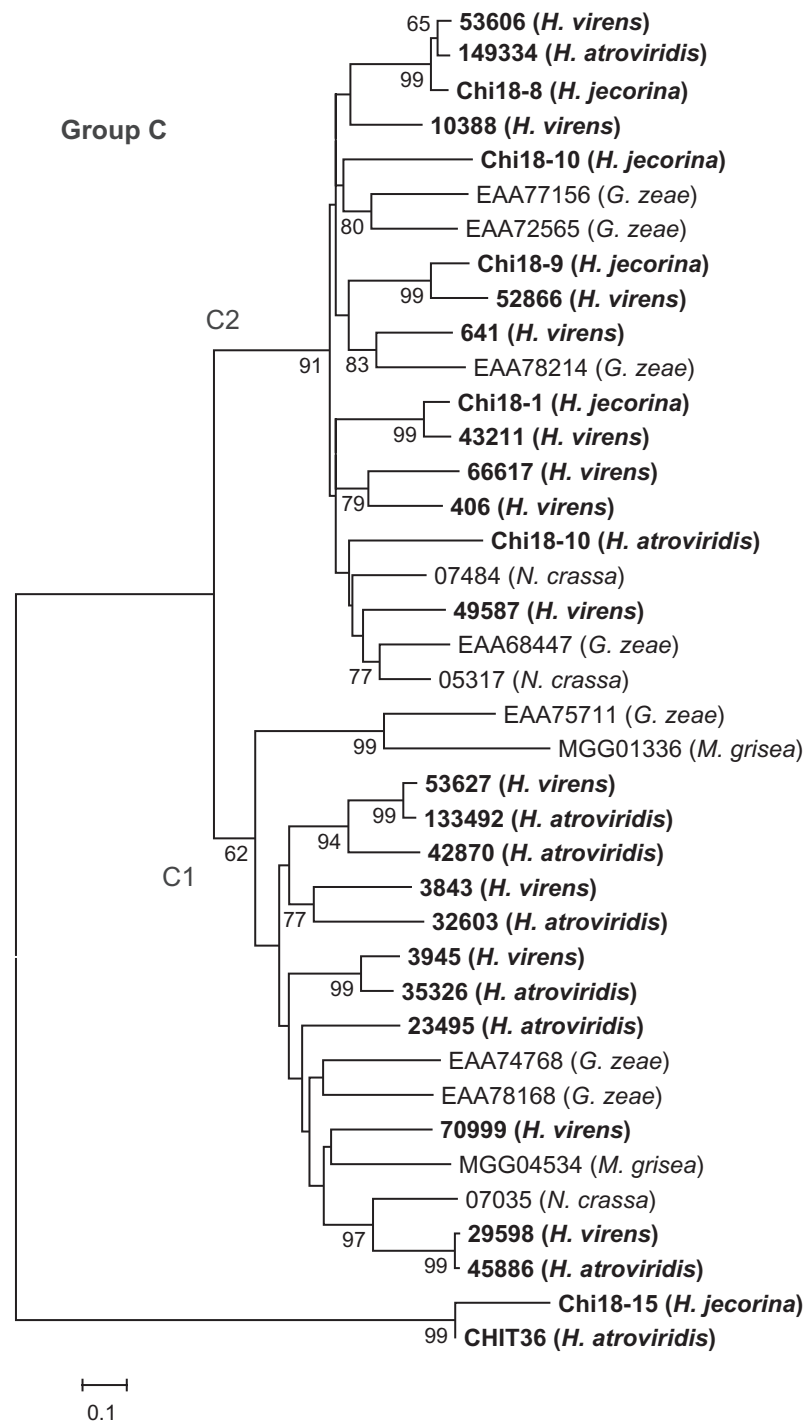
**Figure S1.** Distribution of chitinase gain and loss among fungal lineages. Phylogenetic relationships among the fungal species used in the current study are shown, including divergence dates in millions of years. Circled numbers represent total number of chitinase genes in extant species and estimates of total number of chitinase genes for ancestral species. Boxed taxon names indicates a significant ( $P$ -values  $\leq 0.05$  or Likelihood ratios  $\geq 50$ ) expansion (+), or a significant contraction (-) of the chitinase gene family size.



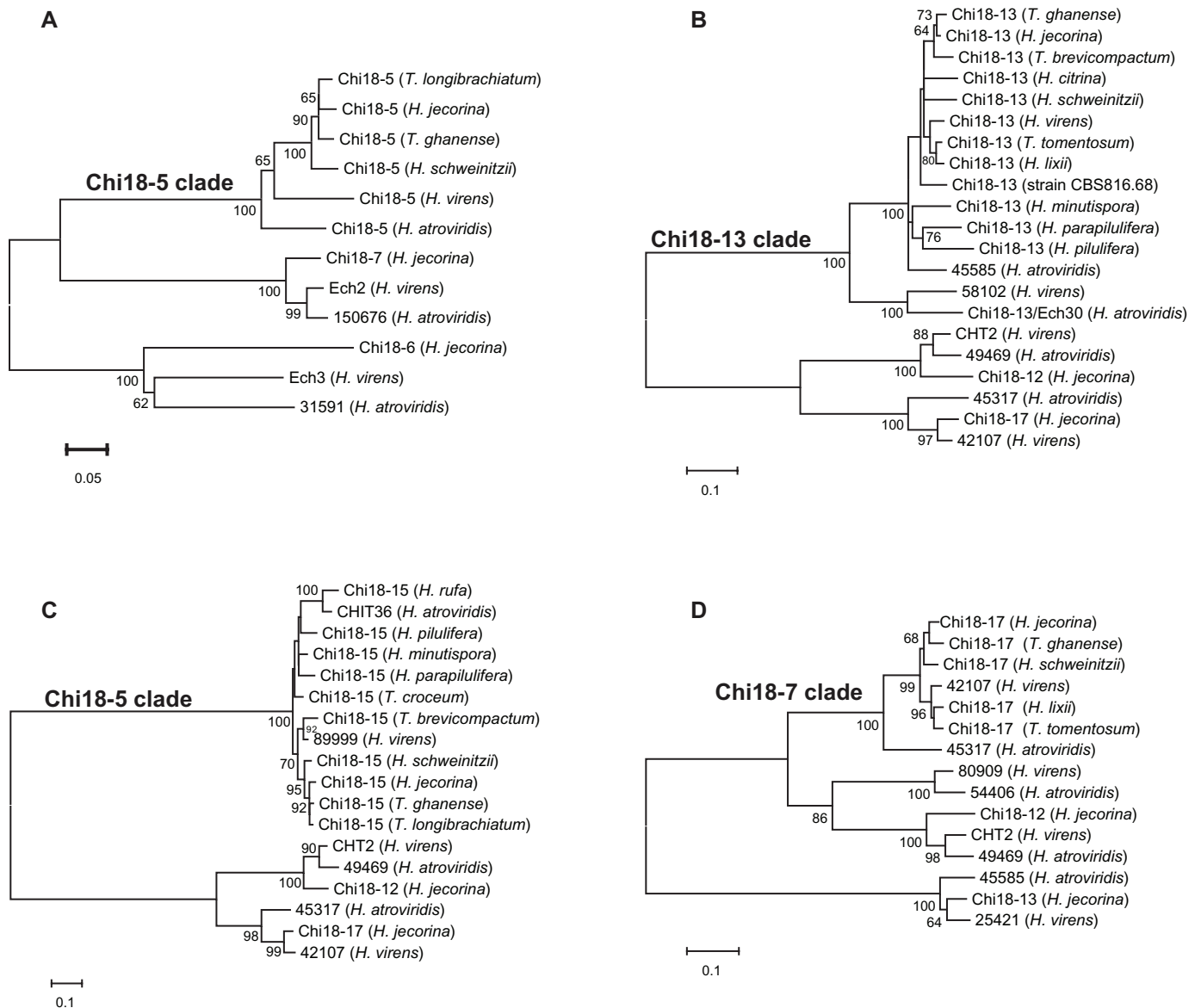
**Figure S2 Group A.** Phylogeny of group A *Trichoderma* chitinases. Analysis was performed using neighbour-joining implemented in MEGA version 4 with the Poisson correction of substitution rates and complete deletion of missing data, based on a Clustal W alignment of chitinase catalytic domain amino acid sequences. Branch support values (bootstrap proportions  $\geq 60$ ) are associated with nodes. The bar marker indicates numbers of amino acid substitutions. Protein identifiers include protein name (if available) or protein ID nos. from the respective genome projects. Subgroup names are indicated.



**Figure S2 Group B.** Phylogeny of group B *Trichoderma* chitinases. Analysis was performed using neighbour-joining implemented in MEGA version 4 with the Poisson correction of substitution rates and complete deletion of missing data, based on a Clustal W alignment of chitinase catalytic domain amino acid sequences. Branch support values (bootstrap proportions  $\geq 60$ ) are associated with nodes. The bar marker indicates numbers of amino acid substitutions. Protein identifiers include protein name (if available) or protein ID nos. from the respective genome projects. Group names are indicated.



**Figure S2 Group C.** Phylogeny of group C Trichoderma chitinases. Analysis was performed using neighbour-joining implemented in MEGA version 4 with the Poisson correction of substitution rates and complete deletion of missing data, based on a Clustal W alignment of chitinase catalytic domain amino acid sequences. Branch support values (bootstrap proportions  $\geq 60$ ) are associated with nodes. The bar marker indicates numbers of amino acid substitutions. Protein identifiers include protein name (if available) or protein ID nos. from the respective genome projects. Group names are indicated.



**Figure S3.** Trichoderma chitinase gene phylogenies. Analyses were performed on chitinases (A) chi18-5, (B) chi18-13, (C) chi18-15 and (D) chi18-17 using neighbour-joining implemented in MEGA version 4 with the Poisson correction of substitution rates and complete deletion of missing data, based on a Clustal W alignment of chitinase amino acid sequences. Branch support values (bootstrap proportions  $\geq 60$ ) are associated with nodes. The bar marker indicates numbers of amino acid substitutions. Protein identifiers include protein name (if available) or protein ID nos. from the respective genome projects.

**Table S1.** Oligonucleotide primers used in the current study.

Gene	Oligonucleotide	Sequence (5'→3')
<i>Chi18-5</i>	Chi18-5.F1	AAGCACTATGCSGATGATT
	Chi18-5.F2	GGTTACmTGTyrkTGCCAAG
	Chi18-5.F3	CAACGTTGCTAsACTTGG
	Chi18-5.F4	TGAAGGAYTGGGGyTTyG
	Chi18-5.F5	GTTCCCGCAArCAAGATT
	Chi18-5.F8	GGTrTCTGGGAYTACAAGG
	Chi18-5.R1	CTGTATATACgTrTkTGCCTATG
	Chi18-5.R2	ArCAGAAGrATCATGTTGG
	Chi18-5.R3	CAYTsGCTCTCTTCTCAAC
	Chi18-5.R4	GwAAGkCTGGCCAATrCCAG
	Chi18-5.R6	GCCTATGTACAGTGGTATATGTG
	<i>Chi18-13</i>	Chi18-13.F1
Chi18-13.F2		GGAGCCACCTTCCGGATT
Chi18-13.F4		ATGwTCTTCAGCAAAGCwCT
Chi18-13.F6		TGGGTCCAGTTyTACAAyAA
Chi18-13.R1		CGGGAACmCATGATGACAC
Chi18-13.R2		CTGrGCyTCCCArAGCAT
Chi18-13.R3		GTTGTAGAACTGrACCCAGATGT
<i>Chi18-15</i>	Chi18-13.R4	GAGCAGACrCCGTTCTTG
	Chi18-15.F2	CwTwmrGAmATCCTACGTTAC
	Chi18-15.F3	TGARwGAATACTACCTTCTCrA
	Chi18-15.F4	ACAAMACGGCTACAACGTGA
	Chi18-15.F5	GGCGTyCTyGCrCAGATG
	Chi18-15.F6	ATCGACATTGACATCGAGAC
	Chi18-15.F7	GTTCCCTAyAryATGCAA
	Chi18-15.F8	TACAGCAyTATGGGAArACG
	Chi18-15.R1	GCCGTTGTGACGATTTTCT
	Chi18-15.R3	CryCsGCwyTCTTCyrGTATCA
	Chi18-15.R4	ACTTTmACAGTrGCrTCCAT
<i>Chi18-17</i>	Chi18-15.R5	CrTCCCAGTTGATwGACCA
	Chi18-15.R6	TGTAATGCTACCACCTGTGA
	Chi18-15.R7a	GTGAyrAmTTTATATACTG
	Chi18-15.R7b	CAGAGCAmyCATAcATGTcG
	Chi18-15.R8	CCATAGTCAAGCCAAAGT
	Chi18-17.F1	ATTCATGTAACCATGTcAG
	Chi18-17.F2	GCsArTyCAGAGCTGTCT
	Chi18-17.F3	AGACvATCATGATGAGTCTG
	Chi18-17.F4	TCATGCTGTGGGATATGG
	Chi18-17.R1	AAGkGAATATwTACAAACAGAA
Chi18-17.R2	CCGTTGTTGTAGAACTGAAC	
Chi18-17.R3	CCAACCATGAAGGArATGT	
Chi18-17.R4	AATCTTGGCCAGGTATCC	



**Table S2.** Residues of *Trichoderma* models corresponding to the catalytic and substrate binding residues of 1FFR (*Serratia marcescens* Chitinase A) structure. Residue numbering refers to catalytic module sequences used for modelling; 15–424 (chi18-5), 30–320 (chi18-13), 25–322 (chi18-15), 28–311 (chi18-17).

Site	1FFR	Chi18-5	Chi18-13	Chi18-15	Chi18-17
Active site	D313	D148	D150	D146	D134
	E315	E150	E152	E148	E136
	F/Y390	Y218	Y209	Y207	Y216
	D311	D146	D148	D144	D132
-5	Y170	Y29	—	—	—
-4	R172	R31	—	—	—
-3	W167	W26	—	—	W23
	T276	T111	V103	—	T101
-2	E473	E295	—	—	—
	W275	W110	A102	S98	A100
	T276	T111	V103	S98	T101
	E473	E295	—	—	—
-1	W539	W357	W279	W281	W298
	E540	E358	E280	D282	S299
	Y163	Y22	Y16	Y13	Y19
	W275	W110	A102	S98	A100
	D313	D148	D150	D146	D134
	E315	E150	E152	E148	E136
	A362	A190	A183	A179	A175
	M388	M216	Q207	Q205	Q214
	F/Y390	Y218	Y209	Y207	Y216
	D391	D219	N210	—	—
	Y444	—	—	—	—
	R446	R274	?	—	—
W539	W357	W279	W281	W298	
+1	W275	W110	A102	S98	A100
	E315	E150	E152	E148	E136
	F316	Y151	—	—	—
	M388	M216	Q207	Q205	Q214
+2	D391	D219	N210	—	—
	R446	R274	?	—	—
	W275	W110	A102	S98	A100
	K369	N197	—	—	—
	D391	D219	N210	—	—
	F396	W224	—	—	—
	Y418	F246	N213	G210	—



**Table S3.** Species and acc. nos. for gene sequences retrieved from GenBank, and used in the current work.

<b>act</b>	<b>GenBank</b>	<b>tef</b>	<b>GenBank</b>	<b>ChiJ</b>	<b>GenBank acc.</b>
<b>Species</b>	<b>acc.nos.</b>	<b>Species</b>	<b>acc.nos.</b>	<b>Species</b>	<b>nos.</b>
<i>Hypocrea flaviconidia</i>	DQ111960	<i>H. andinensis</i>	EU280042	<i>Streptomyces avermitilis</i>	NP_826813
<i>H. lutea</i>	FJ442602	<i>H. crassa</i>	EU280053	<i>S. clavuligerus</i>	ZP_05007692
<i>H. melanomagna</i>	FJ442601	<i>H. koningii</i>	EU280017	<i>S. ghanaensis</i>	ZP_04683727
<i>H. minutispora</i>	DQ111977	<i>H. lixii</i>	EU279994	<i>S. griseus</i>	YP_001828239
<i>H. pachybasioides</i>	DQ111976	<i>H. novaezelandiae</i>	EU280039	<i>S. sp. Mg1</i>	ZP_04999219
<i>H. rufa</i>	DQ333563	<i>H. orientalis</i>	EU280038	<i>S. coelicolor</i>	NP_626743
<i>H. stilbohypoxyli</i>	DQ111967	<i>H. tawa</i>	EU279972		
<i>H. virens</i>	FJ442590	<i>H. virens</i>	EU280065		
<i>H. viridescens</i>	FJ442594	<i>H. lutea</i>	EU280058		
<i>Trichoderma asperellum</i>	EU856272	<i>T. asperellum</i>	EU279961		
<i>T. austrokingii</i>	DQ379011	<i>T. atroviride</i>	EU280024		
<i>T. caribbaeum</i>	DQ328610	<i>T. brevicompactum</i>	EU280061		
<i>T. dingleyae</i>	DQ367718	<i>T. citrinoviride</i>	EU280036		
<i>T. dorotheae</i>	DQ379009	<i>T. cuenispurum</i>	EU280052		
<i>T. erinaceum</i>	DQ323450	<i>T. gamsii</i>	EU280005		
<i>T. evansii</i>	EU856269	<i>T. ghanense</i>	EU280043		
<i>T. hamatum</i>	EU856267	<i>T. helicum</i>	EU280055		
<i>T. koningiopsis</i>	DQ379014	<i>T. koningiopsis</i>	EU280012		
<i>T. lieckfeldtiae</i>	EU856276	<i>T. longipile</i>	EU280051		
<i>T. ovalisporum</i>	DQ328611	<i>T. ovalisporum</i>	EU280004		
<i>T. ovalisporum</i>	DQ328608	<i>T. pleurotica</i>	EU279973		
<i>T. pauculosporum</i>	DQ111957	<i>T. pleurotum</i>	EU279975		
<i>T. petersenii</i>	DQ379013	<i>T. rossicum</i>	EU280066		
<i>T. pubescens</i>	EU856249	<i>T. saturnisporum</i>	EU280044		
<i>T. rogersonii</i>	DQ367716	<i>T. sinensis</i>	EU280041		
<i>T. spirale</i>	FJ442819	<i>T. spirale</i>	EU280050		
<i>T. strigosum</i>	DQ111964	<i>T. tomentosum</i>	EU279971		
<i>T. taiwanense</i>	DQ323455				
<i>T. theobromicola</i>	EU856270				
<i>T. viride</i>	DQ111970				



**Publish with Libertas Academica and every scientist working in your field can read your article**

*"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."*

*"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."*

*"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."*

**Your paper will be:**

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

**<http://www.la-press.com>**