

Supporting Information Appendix

for

Multiple lateral gene transfers and duplications promoted plant parasitism ability in nematodes

Etienne G.J. Danchin^{a,1}, Marie-Noëlle Rosso^a, Paulo Vieira^a, Janice de Almeida-Engler^a, Pedro M. Coutinho^b, Bernard Henrissat^b, Pierre Abad^a.

^aINRA, UMR 1301, Plant / Nematodes Interaction, F-06903 Sophia-Antipolis, France.

^b CNRS, UMR6098, Architecture et Fonction des Macromolécules Biologiques, Universités Aix-Marseille, F-13009, Marseille, France.

¹Corresponding Author

Etienne G.J. Danchin,

e-mail: etienne.danchin@sophia.inra.fr

Address: INRA, UMR 1301, IBSV, Plant / Nematodes Interaction, 400 route des Chappes, BP 167, F-06903 Sophia-Antipolis Cedex, France.

Table of contents :

SI Materials and Methods	pp. 3-5
SI Figure and Table Legends	pp. 6-10
Table S1	pp. 11-12
Figure S1 A-B	pp. 13
Figure S2 A-B	pp. 14-15
Figure S3 A-C	pp. 16-18
Figure S4	pp. 19
Figure S5 A-C	pp. 20-22
Figure S6 A-B	pp. 23-24
Figure S7 A-B	pp. 25-26
Figure S8 A-B	pp. 27-28
Tables S2 A-F	pp. 29-34
Figure S9 A-D	pp. 35
Figure S10 A-D	pp. 36
Figure S11 A-F	pp. 37-68
Table S3	pp. 69
Figure S12	pp. 70
Table S4 A-B	pp. 71-72
Figure S13	pp. 73
Figure S14	pp. 74
Table S5	pp. 75
Figure S15	pp. 76
SI References	pp. 77-79
Dataset S01	pp. 80-82
Dataset S02	pp. 83-87
Dataset S03	pp. 88-90
Dataset S04	pp. 91-96
Dataset S05	pp. 97-99
Dataset S06	pp. 100-101
Dataset S07	pp. 102-104

Supporting Information (SI)

SI Material and Methods

Similarity search in public databases

For each plant cell wall-degrading enzyme family, previously identified in the genome sequence of *M. incognita* (1), candidate homologs were searched in an ensemble of databases using as queries one or several protein sequences from *Meloidogyne incognita*, depending on whether or not they were part of a multigenic family in this species.

At the protein level, blastP (2) search was done against the following databases:

- CAZy database of Carbohydrate-Active enzymes specific catalytic and appended modules including all species currently present (3).
- Predicted peptides from *Heterodera glycines* 3X coverage-assembly (patent: US-2007/0271630-A1).
- Predicted proteins from *Meloidogyne hapla* 10X coverage assembly (4), www.hapla.org.
- Non redundant (nr) protein database at the NCBI.

At the nucleotide level, TblastN search was done against the following databases:

- Nematode.net collection of EST clusters (specifically against clade IVb and *Xiphinema index*) (5).
- Nematodes.org collection of EST clusters (specifically against clade IV and *Xiphinema index*) (6).
- *Heterodera glycines* genomic DNA contigs from the 3X coverage-assembly (patent: US-2007/0271630-A1)
- *Globodera pallida* draft genome assembly (October 2009 version) at the Sanger Centre web portal (<http://www.sanger.ac.uk/sequencing/Globodera/pallida/>).
- EST_others database at the NCBI.

Sequence retrieval

All sequences that aligned at least on 60% of the length of the considered catalytic or active module at the protein level with an e-value not higher than 0.01 were kept for multiple sequence alignment, with a limit of 200 sequences kept for phylogenetic analysis. This max number of 200 sequences systematically allowed including a substantial number of species other than plant-parasitic nematodes in phylogenetic analyses. Nucleic sequences from EST contig databases were translated into protein sequences. Genome assembly contigs from *H. glycines* and *G. pallida* genome that returned significant matches were retrieved for gene prediction. Genes and encoded proteins were predicted using the online version of FGENESH+ (www.softberry.com). Full length proteins from *M. incognita* were used as templates for similarity based prediction and *Caenorhabditis elegans* was chosen for the HMM model. Sequences of a same species that were 100% identical to one another or entirely included in a longer one were eliminated to remove redundancy.

Multiple sequence alignment

Collections of sequences for each cell wall degradation family were aligned using the program MUSCLE (7, 8) with default parameters. Multiple sequence alignments (MSA) were manually examined using JALVIEW(9). Sequences that contained substantive gap numbers and lengths in otherwise high quality portions of alignment were removed from the analysis in order to keep as much as possible informative sites for phylogenetic reconstruction. MSA were systematically recalculated whenever a sequence was removed, edited or appended in the original alignments.

Phylogenetic reconstruction

We performed phylogenetic analyzes using two different approaches, Bayesian estimation and bootstrapped Maximum Likelihood. Bayesian phylogenetic reconstructions were done using the program mrBayes (10) with a mixture of models, an estimated gamma distribution of rates of evolution and an estimation of the proportion of invariable sites. By default 100,000 generations were run for each phylogeny. In case the average standard deviation of split frequencies was not inferior to 0.01 after 100,000 generations, additional generations were launched until

congruence was reached (<0.01). Consensus trees and statistics were obtained after systematically “burning” 25% of generated trees. Posterior probability support values are reported for each node. To obtain support from a second independent method, we also performed phylogenetic analyses using Maximum Likelihood (ML) estimation with the RAxML software (11). We systematically ran 100 bootstrap replicates followed by a ML search for the best-scoring tree. For each analyzed protein family we selected the model of amino acids evolution that returned the best posterior probability score in corresponding Bayesian phylogenetic analysis. We used a model with four categories of estimated gamma rates of evolution as well as an estimate of the proportion of invariable sites.

Topologies obtained via Bayesian estimation were selected as reference topologies and we reported bootstrap support values from the best-scoring ML trees on corresponding Bayesian topologies. In case a subtree observed in Bayesian topologies was not found in ML topologies, an asterisk (*) was put at the corresponding node. ML bootstrap values were not indicated on terminal nodes as they were systematically higher than 90 and would have prevented clear visualization of terminal nodes. Overall, topologies obtained by the two methods were very similar, only a proportion of support values were significantly different.

Phylogenetic analyses were run on a Linux cluster of 8 CPUs with 16 Go of RAM. Trees were annotated using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). Phylogenetic trees for each plant cell wall-modifying protein family are available in figures S2-S8 and a list of sequences included in the respective phylogenetic analyses are available in Datasets S01-S07. Statistics of the different phylogenetic reconstructions, including the model of evolution chosen, the estimated values of the alpha parameter and of the proportion of invariable sites are summarized in Table S5.

Statistical testing of alternative topologies

For all phylogenetic trees presented in Figures 1 and 2 of the main manuscript we generated constrained alternative trees in which all eukaryotic homologs were grouped in a monophyletic cluster using the program RAxML. We used exactly the same procedure and models as for non-constrained trees and performed an ML analysis with 100 bootstrap replicates. For each family we then selected the best-scoring non-constrained ML tree and a random set of 10 topologies from the 100 bootstrap replicates as well as the best-scoring constrained ML tree and 10 randomly selected topologies from the 100 bootstrap replicates. Hence the 11 first topologies correspond to non-constrained trees and the 11 other topologies correspond to constrained trees. The first

and twelfth topologies correspond to the best-scoring constrained and non-constrained topologies, respectively. We next calculated per-trees log-likelihoods of site patterns using RAxML for the sets of 22 tree topologies. We then used the suite of programs implemented in the CONSEL software (12) to evaluate the different topologies and select the most likely tree. We systematically reported in Tables S2A-S2F p-values for the different statistical tests obtained by CONSEL and used the approximately unbiased (au) test as the main criterion for comparison of topologies and tree selection.

GC content and codon usage

We used the program geece from the EMBOSS software package (13) to calculate the GC content of every CDS from predicted plant cell wall-modification (PCWM) gene as well as for all the rest of CDS predicted from the *M. incognita* genome. The first dataset includes all PCWM CDS (including expansin-like proteins and incomplete gene models) and comprises 85 sequences (Table S3). The second dataset is composed of all the 20,365 CDS (including splice variants) predicted from the *M. incognita* genome (Table S3). The distribution of the GC content for the two datasets has been plotted in two aligned graphs for comparison purpose (Figure S12, Table S3). We also analyzed and compared the codon usage for the same two datasets, using the program cusp from the EMBOSS package (Table S4A). Codons that were found to be more than five percent differentially used between the two datasets are listed in Table S4B.

Genomic clusters of genes encoding PCW active proteins

Positions of all genes encoding plant cell wall-modifying and degrading proteins included in this analysis were retrieved in the genome assembly of *M. incognita*. We generated a map plotting the different genes on the 17 scaffolds that contained at least two plant cell wall-active genes (Figure S13). The map was generated using the R statistical system (<http://www.R-project.org>) with the package R/qtl (14).

Determination of exon / intron structures

Exon / intron structures of genes encoding plant cell wall-modifying proteins were determined using two different approaches. For species with a publicly available genome sequence (i.e. *M. incognita*, *M. hapla*, *H. glycines*, *G. pallida*), exon / intron structures were predicted based on alignment of corresponding protein sequences with genome assemblies, using the program SCIPIO (15) and output produced by the associated online tool WEBSCIPIO (16). For other species, protein sequences were aligned to corresponding genomic sequences (when available) using the program Wise 2 (<http://www.ebi.ac.uk/Tools/Wise2/index.html>). Positions of

introns were reported on the protein sequences by inserting the “XX” characters at the junction between two consecutive exons. We generated multiple alignments for each of the six plant cell wall-modifying protein families to determine conservation of intron positions between species and clades (Figure S11). Only sequences that had no obvious gene model errors (i.e. partial sequences or artificial fusions) were kept for this analysis.

Antibody and serum production for immunolocalization

A serum directed against the CBM2 module of *Meloidogyne incognita* ENG1 protein was produced by rabbit immunization with synthetic peptides (Eurogentec Polyclonal Antibody Production, France). The selected peptide (NGPATPPQIKVLGDGKC) also perfectly matches the CBM2 modules of *Meloidogyne incognita* ENG1a and ENG3 proteins and shares significant similarity with the CBM2 module of *Meloidogyne incognita* expansin-like proteins (Figure S15). In order to prevent background noise, this peptide was selected based on the lack of similarity with other nematode proteins or with plant proteins. In addition, the selected peptide is localized on hydrophilic regions of the proteins with good accessibility to solvents as predicted by the Porter and PaleAle servers (<http://distill.ucd.ie/paleale/>; (17)).

Immunocytochemical analysis in tomato roots

Roots of *Lycopersicon esculentum* cv. St. Pierre tomato and *Arabidopsis thaliana* plants were inoculated with *M. incognita*. Dissected roots and galls, taken at different time points after inoculation (21 and 55 days after inoculation), were fixed with 4% formaldehyde in 50 mM Pipes buffer (pH 6.9). Samples were then dehydrated, embedded in butyl-methylmethacrylate (18), sectioned (5µm), and affixed to Poly-L-Lysine-coated slides. Slides holding feeding sites were immunolabelled following the methodology described in (19). Primary (anti-CBM2) and secondary (Alexa 488 goat anti-rabbit IgG, Molecular Probes) antibodies were diluted 1:200 fold in blocking solution (1% bovine serum albumin in 50 mM Pipes, pH 6.9). Slides were washed in Pipes for 30 min at RT and nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI; 1µg/ml in PBS) for 5 min. Control sections were incubated with pre-immune serum or in the absence of primary antibody. Mounted slides (Prolong antifade kit; Invitrogen Molecular Probes, Eugene, OR, U.S.A.) were observed with a microscope equipped for epifluorescence microscopy and differential interference contrast optics (Axioplan 2, Zeiss) and images were taken with an AxioCam digital camera (Zeiss).

SI figure and table legends

Fig. S1. Overview of the phylogeny of Nematoda and detail on Tylenchina lineage.

A schematic representation of the phylogeny of the Nematoda phylum is represented on panel (A). The simplified tree topology is adapted from recent phylogenies based on small subunit ribosomal DNA sequences (20, 21). Position of the four clades (I to IV) as defined by Blaxter *et al.* in 1998 (22) are indicated at corresponding nodes. Example species are given for each phylum in brackets and those in which plant-parasitic species are found are highlighted in green. Species for which a genome sequence is publicly available are represented with an asterisk (*) in superscript. Species abbreviated names stand for *Meloidogyne incognita*, *Meloidogyne hapla*, *Globodera pallida*, *Aphelenchus avenae*, *Zeldia punctata*, *Panagrellus redivivus*, *Strongyloides ratti*, *Steinernema carpocapsae*, *Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Haemonchus contortus*, *Pristionchus pacificus*, *Brugia malayi*, *Ascaris suum*, *Rhigonema thysanophora*, *Oxyuris equi*, *Plectus aquatilis*, *Monhystera riemannii*, *Chromadorita cf. leuckarti*, *Xiphinema index*, *Mononchus truncatus*, *Mermis nigrescens*, *Trichinella spiralis*, *Trichodorus cylindricus*, *Enoplus meridionalis* from top to bottom and from left to right. A detailed view of the Tylenchina lineage is represented on panel (B). The tree topology is based on published reports (21, 23). Commonly used names are given for the two most economically relevant lineages of plant parasites, Meloidogynidae (root knot nematodes) and Heteroderidae (cyst nematodes). Example species are given for each lineage. Species abbreviated names stand for *Meloidogyne incognita*, *Meloidogyne hapla*, *Pratylenchus coffeae*, *Nacobbus aberrans*, *Heterodera glycines*, *Globodera pallida*, *Rotylenchus robustus*, *Radopholus similis*, *Paratylenchus microdorus*, *Ditylenchus dipsaci*, *Sphaerularia bombi*, *Aphelenchus avenae* from top to bottom and from left to right.

Table S1. Summary of functional data and phylogenetic analyses of plant cell wall-degrading and modifying proteins in PPN.

For each protein family involved in modification or degradation of the plant cell wall a series of functional and evolutionary information is summarized and fully referenced. Functional data includes enzymatic activity, transcript and protein localization and effect of inactivation (where applicable). Evolutionary information includes presence pattern in plant-parasitic nematodes, summary of phylogenetic trees topologies, lists of species possessing proteins most closely related to those of nematodes as well

as duplication pattern and degree of conservation of exon / intron structures.

Fig. S2. Phylogenetic tree of RKN GH28 polygalacturonases.

(A) This GH28 tree topology was obtained from the Bayesian analysis and rooted at the midpoint. Phylogenetic groups were color-coded as follows; plant-parasitic nematodes in green, Plants in blue and bacteria in red. Posterior probability support values are indicated at each node (top value) as well as corresponding bootstrap values (bottom value). A list of sequences and species included in this analysis is available in Dataset S01. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (where applicable). (B) The same phylogenetic tree, using the same annotation but represented unrooted in a radial form, this unrooted tree topology was used to represent Figure 1A of the main manuscript. All phylogenetic groups supported by posterior probability values >0.70 and grouping species from a single kingdom were collapsed for ease of visualization. Only phylogenetic groups containing nematodes and their closest bacterial relative are expanded.

Fig. S3. Phylogenetic tree of PL3 Pectate Lyases.

(A) Bayesian unrooted tree topology of PL3 family. Posterior probability support values are indicated at each node (top value) as well as corresponding bootstrap values (bottom value) from the ML analysis. Branches emerging from deep nodes with high posterior probability values were collapsed for ease of visualization. Resulting phylogenetic groups were color-coded as follows; plant-parasitic nematodes in green, fungi in dark blue, oomycetes in light blue and bacteria in red. The fully expanded phylogenetic tree with the same color-coding is available in fig. S3C and a list of sequences and species included for that analysis is available as Dataset S02. (B) Detailed phylogenetic tree including only nematode PL3 Pectate Lyases and their closest relatives (thus excluding fungi and oomycetes). Posterior probability support values are indicated at each node as well as the corresponding bootstrap values obtained from the ML analysis. The tree was presented unrooted and phylogenetic groups were color-coded using the same model as for fig. S3A. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (when applicable). A list of species names corresponding to abbreviated names is available in Dataset S02. Nodes of interest are highlighted with a red circle and annotated / numbered according to the following comments: (i) groups PL3s from Aphelenchoidea, (ii) groups a cluster of PL3s from actinobacteria, (iii) groups PL3s from cyst nematodes and parts of those from root knot nematodes with another cluster of actinobacteria, (iv) groups the rest of

PL3s from root knot nematodes. (C) This Bayesian tree represents the same topology than fig. S3A but with full details of taxa included in each cluster, using the same color code. The tree was rooted at the midpoint and represented as a phylogram. Posterior probability values are indicated at each node. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (where applicable).

Fig. S4. Phylogeny of GH43 candidate arabinanases.

This Bayesian tree topology was used to produce unrooted tree in Figure 1C. Posterior probability support values are indicated at each node as well as the corresponding bootstrap values obtained from the ML analysis. The tree is presented as a phylogram rooted at the midpoint and phylogenetic groups are colored using the following model: plant-parasitic nematodes in green, bacteria in red, fungi in dark blue and Oomycetes in light blue. Leaves of the tree are labeled as follows; species abbreviated name, accession number. A list of species names corresponding to abbreviated names is available in Dataset S03.

Fig. S5. Phylogenetic tree of GH5 cellulases.

(A) Unrooted Bayesian topology of GH5 cellulases. Posterior probability support values are given at each node as well as the corresponding bootstrap values obtained from the ML analysis. Branches emerging from deep nodes with high posterior probability values were collapsed for ease of visualization. Resulting phylogenetic groups were color-coded as follows; plant-parasitic nematodes in green, other Metazoa in blue, bacteria in red and symbiotic protists in light orange. The fully expanded phylogenetic tree with the same color-coding is available in fig. S5C and a list of sequences and species included for that analysis is available as Dataset S04. (B) Detailed tree of plant-parasitic nematode GH5 cellulases and their closest relatives (thus, excluding the most distant bacterial groups from tree S5A). This tree topology was used to produce unrooted tree in Figure 2A. Posterior probability support values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. The tree was rooted at the midpoint and phylogenetic groups were color-coded using the same model as for fig. S5A. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (where applicable). A list of species names corresponding to abbreviated names is available in Dataset S04. Nematode cellulases that possess a CBM2 module are highlighted in red and their last common ancestral node is highlighted by a red circle. (C) Expanded phylogenetic tree of GH5 cellulases. This tree represents the same topology than fig. S5A but with full details of taxa included in each cluster using the same color

code. The tree was rooted at the midpoint and represented as a phylogram. Posterior probability support values are indicated at each node. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (where applicable).

Fig. S6. Phylogenetic tree of GH5 xylanases.

(A) This Bayesian tree topology includes GH5 xylanases and the related GH30 proteins. Posterior probability support values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. The tree was rooted at the midpoint and phylogenetic groups were color-coded as follows; plant-parasitic nematodes in green, fungi in blue and bacteria in red. Branches corresponding to GH30 enzymes were colored in black. A list of sequences and species included for this analysis is available as Dataset S05. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (when applicable). (B) Phylogenetic tree of GH5 xylanases including sequences translated from ESTs and excluding GH30 proteins that formed a distinct outlier in fig. S6A. This Bayesian tree topology was used to produce unrooted topology in Figure 2B. Posterior probability support values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. The tree was rooted at the midpoint and represented as a phylogram. Phylogenetic groups were color-coded using the same convention as Fig. S6A. Leaves of the tree were labeled as follows; species abbreviated name, accession number, usual name (when applicable).

Fig. S7. Phylogenetic tree of EXPN expansin-like proteins.

(A) This unrooted bayesian tree topology includes expansin-like proteins from nematodes, bacteria, fungi, amoeba and oomycetes. Posterior probability support values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. Phylogenetic groups were colored using the following model: plant-parasitic nematodes in green, bacteria in red, fungi in dark blue, amoeba in violet and oomycetes in light blue. Leaves of the tree were labeled as follows; species abbreviated name, accession number. A list of species names corresponding to abbreviated names is available in Dataset S06. Nodes of interest are highlighted with a red circle. (B) Detailed phylogenetic tree of expansin-like proteins, excluding sequences from oomycetes and from the most distant fungi. This tree topology was used to produce unrooted topology in Figure 2C. Posterior probability support values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. According to the

topology presented in fig. S7A; fungi were selected as outgroup. Phylogenetic groups were colored using the following model: plant-parasitic nematodes in green, bacteria in red, fungi in dark blue and amoeba in violet. Leaves of the tree were labeled as follows; species abbreviated name, accession number. Nodes of interest are highlighted with a red circle. Nematode expansin like proteins that bear a CBM2 module are highlighted in red. A list of sequences and species included for that analysis is available as Dataset S07.

Fig. S8. Phylogenetic tree of CBM2 modules.

(A) Bayesian tree topology of nematode and bacterial CBM2 modules. Posterior probability values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. The tree was rooted at the midpoint and phylogenetic groups were color-coded as follows; plant-parasitic nematodes in green, bacteria in red. Leaves of the tree were labeled as follows; species abbreviated name, accession number, associated module. (B) Tree topology obtained from a reduced dataset containing nematode CBM2 modules and their two most closely related bacterial orthologs. Posterior probability values as well as the corresponding bootstrap values obtained from the ML analysis are indicated for each node. The three bacterial CBM2 sequences were chosen as outgroups.

Table S2. Statistics of comparison of constrained and non-constrained tree topologies.

We systematically reported in tables S2A-S2F p-values for the different statistical tests obtained by CONSEL (12) for comparison of constrained and unconstrained trees. We used the approximately unbiased (au) test as the main criterion for comparison of topologies and tree selection. Abbreviations are as follows “au” for approximately unbiased test, “np” and “bp” for bootstrap probabilities of the selection, “pp” for Bayesian posterior probability, “kh” for Kishino-Hasegawa test, “sh” for Shimodaira-Hasegawa test, “wkh” for weighted Kishino-Hasegawa test, “wsh” for weighted Shimodaira-Hasegawa test. The best constrained and non-constrained ML trees are highlighted in bold. (A) GH28 polygalacturonases. The constraint consisted in grouping nematode and plant GH28 homologs in a monophyletic group. (B) PL3 pectate lyases. The constraint consisted in grouping nematode, oomycete and fungal PL3 homologs in a monophyletic group. (C) GH43 candidate arabinanases. The constraint consisted in grouping nematode, oomycete and fungal GH43 homologs in a monophyletic group. (D) GH5 cellulases. The constraint consisted in grouping nematode, insect and symbiotic protists GH5 cellulase homologs in a monophyletic group. (E) GH5 xylanases. The constraint consisted in grouping nematode, and fungal GH5 xylanases homologs in a monophyletic group.

(F) Expansin-like (EXPN) proteins. The constraint consisted in grouping nematode, amoeba and fungal EXPN candidate expansin homologs in a monophyletic group.

Fig. S9. Immunodetection of CBM2-bearing proteins in migratory *Meloidogyne incognita* during parasitism of *Arabidopsis thaliana* roots.

(A and B) Root section containing a migratory nematode, showing the localization of CBM2-bearing proteins in the subventral glands of the nematode, and secreted into root tissue around the head of the nematode (orange arrow). Left upper panel corresponds to overlay of images of CBM2-bearing proteins (green) and DAPI stained nuclei (blue). Right upper panel corresponds to overlay of images of CBM2-bearing proteins (green), DAPI stained nuclei (blue) and differential interference contrast (gray) to better visualize plant and nematode tissues. (C and D) Root section containing migratory nematodes incubated with the pre-immune serum as a control, showing no signal (green). N stands for nematode, m for metacorpus, svg for subventral glands. Scale bar = 10 μ m.

Fig. S10. Controls of late parasitic stage immunolocalization assays in tomato roots.

Results of control immunodetection with the pre-immuno sera used in place of the primary antibody in roots of tomato (*Lycopersicon esculentum*) infected with *Meloidogyne incognita*. The two top panels (A and B) show eggs (e, white arrow) in ovary region of an *M. incognita* adult female. The two bottom panels (C and D) highlight the vulva region of an adult female pointed with a white arrow (v). As for Fig. 3, N stands for Nematode, G for giant cell, gm for gelatinous matrix. Scale bar = 10 μ m.

Table S3. GC content distribution of plant cell wall modification genes and the rest of protein-coding genes in *M. incognita*.

The first column lists the observed range of GC percent (from 16 to 35 %). For a given GC percent, the number of CDS and of cell wall-modifying CDS in the *M. incognita* genome are shown in columns two and three, respectively. For the ensemble of protein-coding CDS, the mean GC percent is 36.26 % (median = 36.00 %). For plant cell wall-modifying CDS, the mean GC percent is 36.33 % (median = 36.50 %). Aligned graphs of the two distributions are shown in fig. S12.

Fig. S11. Position of introns in genes encoding PCW-modification proteins.

For each family of plant cell wall-modifying protein, a multiple sequence alignment displays position of introns marked by “XX” characters and highlighted by a red background color.

The rest of positions in the multiple sequence alignment are background-colored using a gradient of blue (from light blue to dark blue) according to the degree of conservation of a given amino-acid at a given position. (A) Position of introns in the GH28 family. This family is only present in RKN thus conservation between different lineages of nematodes cannot be assessed, though all intron positions are shared between *Meloidogyne hapla* and *Meloidogyne incognita*. (B) Position of introns in the PL3 family. Conservation of introns in cyst nematodes (CN) and / or root knot nematodes (RKN) and / or Aphelenchoideae (Aph) is indicated by a legend pointing to the corresponding position with a red arrow. Two intron positions are shared between CN and RKN, one intron position is shared between the three nematode lineages. (C) Position of introns in the GH43 family. This family has been identified only in RKN and CN and data about intron / exon structure for a full length protein is only available for *Globodera pallida* (CN), *Meloidogyne incognita* and *Meloidogyne hapla* (RKN). Conservation of introns in CN and / or RKN is indicated by a legend pointing to the considered position with a red arrow. One intron position is shared between RKN and CN. (D) Position of introns in the GH5_cel family. Conservation of introns in CN and / or RKN and / or Anguinidae (Ang) and / or Radopholinae (Rad) is indicated by a legend pointing to the corresponding position with a red arrow. Two intron positions are shared between CN, Ang and Rad, two positions are shared between CN and RKN, one position is shared between Ang and Rad, one position is shared between CN and Rad, one position is shared between CN, RKN and Rad and one position is shared by all the four lineages. (E) Position of introns in the GH5_xyl family. This family is only found thus far in RKN and Radopholinae (Rad). Conservation of introns in RKN and / or (Rad) is indicated by a legend pointing to the corresponding position with a red arrow. All the three intron positions are shared by RKN and Rad. (F) Position of introns in the EXPN family. Conservation of introns in CN and / or RKN and / or Aphelenchoideae (Aph) is indicated by a legend pointing to the corresponding position with a red arrow. Three intron positions are shared between CN and RKN, one intron position is shared between the three nematode lineages.

Fig. S12. Compared distributions of GC content in plant cell wall modification genes and the rest of *M. incognita* genes.

The top graph displays the distribution of abundance of CDSs per given percentage of GC in *M. incognita*. The x axis represents the range of GC percents (from 16 % to 65%). The y axis represents abundance in terms of number of *M. incognita* CDS. The bottom graph displays the distribution of abundance of cell wall-modifying CDSs per given percentage

of GC in *M. incognita*. The x axis was set to the same scale as for the top graph for comparison purpose though GC percent for cell wall modification genes range from 26% to 42%. The y axis represents abundance in terms of number of *M. incognita* plant cell wall-modifying CDS.

Table S4. Codon usage comparison in plant cell wall-modifying genes and the rest of *M. incognita* genes.

(A) This table reports differences in codon usage between plant cell wall-modifying genes and the rest of protein-coding genes in *M. incognita*. The first column lists the codons and the corresponding amino acid is shown on the second column. Columns 3-5 represent respectively, the fraction of codons used for a given encoded amino-acid, the frequency of this codon and the number of codons in CDS of plant cell wall-modifying genes. Columns 6-8 represent respectively, the fraction of codons used for a given encoded amino-acid, the frequency of this codon and the number of codons in CDS of all *M. incognita* protein-coding genes. The last column represents the percentage of difference in codon usage for a given codon. Codons that were more than 5% differentially used between plant cell wall-modifying genes and the rest of protein-coding genes are represented in red and were reported on Table S4B for ease of visualization. (B) Codons more than 5% differentially used between plant cell wall-modifying genes and the rest of *M. incognita* genes. The first column shows the codons, the second represents the encoded amino acids (including stop). The third and fourth columns represent the codon usage in percent in CDS encoding plant cell wall-modifying enzymes and CDS from the rest of the protein coding genes, respectively. Stop codons are the more differentially used with more than 14 and 19 % difference in usage for two of them.

Fig. S13. Genomic clusters of plant cell wall-degradation and modification genes.

A map representing *M. incognita* contigs that bear at least two genes encoding plant cell wall degradation or modification genes is plotted. Genes encoding GH5 cellulases are represented in red, Xylanases in orange, cellulose-binding protein (CBP) in yellow, expansin-like proteins in blue and pectate lyases in green. Out of the 17 corresponding plant cell wall-active gene clusters, only two are constituted of genes from different families. These exceptions consist of an expansin-like gene in the vicinity of a xylanases on contig 65 as well as a candidate Cellulose-Binding Protein (CBP) in proximity of another xylanases on contig 317. The majority of plant cell wall-active gene clusters are family-specific clusters probably resulting from cis-duplications. The phylogenetic analyses reveal that duplications that gave rise to these gene clusters occurred before the separation of *M. incognita* and

M. hapla lineages as, generally, no species-specific phylogenetic groups could be identified. The topologies also indicate that the separation of the two pseudo-allelic regions in *M. incognita* is more recent than the cis-duplication of genes in the considered region.

Fig. S14. Taxonomic relations of bacteria representing potential source of lateral gene transfers.

This schematic tree represents the taxonomic relations between bacteria that can be considered as potential source of lateral gene transfers for genes encoding plant cell wall-modifying proteins in plant-parasitic nematodes. All bacteria that turned out to have the genes most closely related to those of plant-parasitic nematodes are represented in this tree. The topology of the tree and the different bacterial clades' names were deduced from the NCBI's taxonomy resource

(<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html>). For each bacterium, the simplified taxonomic classification is shown as well as the gene family(ies) for which they represent a potential donor. Branches of the tree are colored according to bacterial clades. Species indicated by an asterisk (*) are phytopathogens. Species indicated by a dagger sign (†) are involved in symbiotic interactions with plants.

Table S5. Statistical characteristics, evolutionary models and parameters of phylogenetic analyses.

The different parameters, statistical results and models used for phylogenetic analyses presented in Figs. S2-S8 are summarized in this table. For each family of plant cell wall-degrading or modifying protein, we indicated the corresponding figures (phylogenetic trees), the values of the alpha parameter of the gamma distribution, the evaluation of the proportion of invariable positions, the best evolutionary model according to MrBayes analyses associated with its posterior probability value (and thus the selected for maximum likelihood analyses), the number of generations or bootstrap replicates run and the method used. Interestingly the WAG model turned to systematically be the evolutionary model with the best posterior probability value in Bayesians analyses. Thus the WAG model was selected for all ML analyses.

Fig. S15. Multiple sequence alignment of CBM2 modules found in *M. incognita*.

This alignment highlights the peptide chosen to produce antibodies targeting CBM2 modules and its degree of conservation in CBM2-bearing proteins in *M. incognita*. The region corresponding to the selected peptide (NGPATPPQIKVLGDGKC) is boxed in black and sequences

exactly matching this peptide are boxed in red. In the region corresponding to the selected peptide amino-acids are colored from light blue to dark blue, according to their degree of conservation. Sequences are labeled with their accession number followed by their assignment to a family with the following suffixes: _GH5 for GH5 cellulase, _CBP for cellulose-binding protein, _EXPN for expansin-like protein.

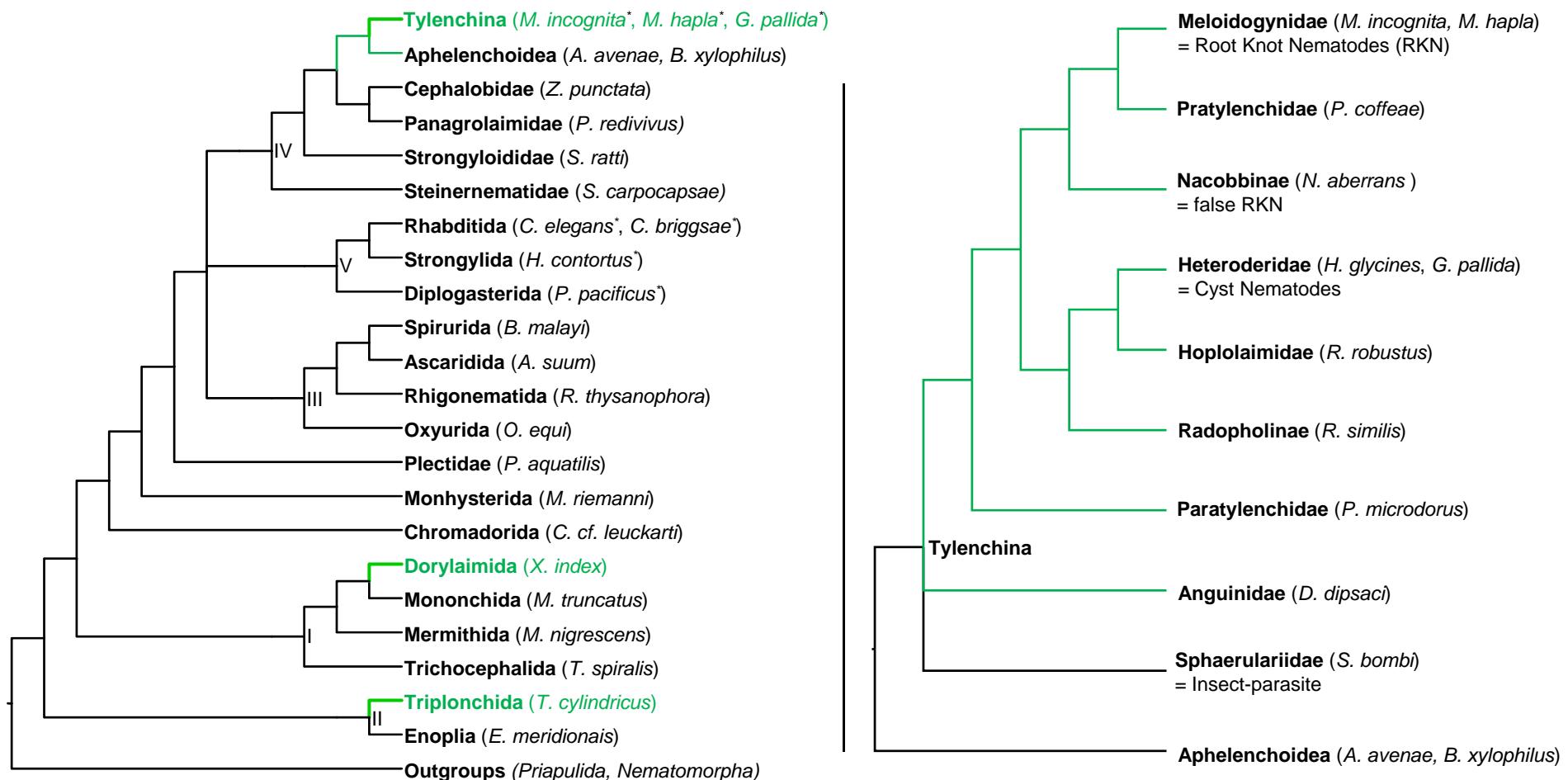
Table S1. Summary of functional data and phylogenetic analyses of plant cell wall-degrading and modifying proteins in plant-parasitic nematodes.

Family	Activity	Presence in PPN	Effect of inactivation	Presence in secretions	Localization / expression in secretory glands	Topology	Closest relative	Duplication Pattern	Exon / Intron Structure
GH28	Polygalacturonase (E.C: 3.2.1.15)	<i>Ditylenchus dipsaci</i> ? (24, 25)			Activity has been claimed but the gene has never been cloned	RKN GH28s grouped with <i>Ralstonia</i> in a single cluster, <i>Ralstonia</i> GH28s splits the RKN cluster in two parts.	<i>Ralstonia solanacearum</i> (Proteobacteria)	Not possible to state precisely due to the absence in CN. Duplications started before separation of <i>M. hapla</i> and <i>M. incognita</i> and continued in <i>M. incognita</i>	Only present in RKN, intron positions conserved.
		<i>Meloidogyne arenaria</i> RKN							
		<i>Meloidogyne chitwoodii</i> RKN							
		<i>Meloidogyne hapla</i> RKN (26)							
		<i>Meloidogyne incognita</i> * RKN (27).	No effect measurable because of silencing duration (28)		In esophageal gland cells of infective juveniles				
		<i>Meloidogyne javanica</i> RKN							
PL3	Pectate Lyase (E.C.: 4.2.2.2)	<i>Aphelenchus avenae</i> (29)			In esophageal gland cells of juveniles and adults	A cluster groups PPN PL3s with Bacteria. PPN clusters are split by two Bacterial clusters	Actinomycetales: <i>Clavibacter michiganensis</i> , <i>Frankia</i> sp., <i>Actinosynnema mirum</i> , <i>Cellulomonas flavigena</i> , <i>Jonesia denitrificans</i> , <i>Streptomyces avermitilis</i> , <i>Streptomyces coelicolor</i>	A few duplications started before the separation of CN and RKN, most occurred independently after the separation of the two lineages but before the sub-differentiation of species.	Two intron positions are shared between CN and RKN, one intron position is shared between the three nematode lineages (RKN, CN, Bursaphelenchus).
		<i>Bursaphelenchus mucronatus</i> (30)							
		<i>Bursaphelenchus xylophilus</i> (30)			In esophageal gland cells				
		<i>Globodera pallida</i> CN							
		<i>Globodera rostochiensis</i> * CN (31)			In subventral esophageal gland cells				
		<i>Globodera tabacum</i> CN							
		<i>Heterodera glycines</i> CN (32, 33)			In subventral gland cells				
		<i>Heterodera schachtii</i> * CN (34)	Significant impairment of parasitism success						
		<i>Meloidogyne arenaria</i> RKN							
		<i>Meloidogyne chitwoodii</i> RKN							
		<i>Meloidogyne hapla</i> RKN							
		<i>Meloidogyne incognita</i> RKN (35)			In subventral gland cells (strong at the pre-parasitic and early parasitic second-stage juveniles, and not detectable at the late parasitic stages of the nematodes)				
		<i>Meloidogyne javanica</i> * RKN (36)		Y	In esophageal gland cells				
GH43	Putative arabinanase	<i>Globodera pallida</i> CN				Monophyly of PPN	Actinomycetales: <i>Sanguibacter keddieii</i> , <i>Solibacter usitatus</i> , <i>Cellulomonas flavigena</i> , <i>Kineococcus radiotolerans</i>	Not possible to state because not present as multigenic family	One intron position shared between RKN and CN.
		<i>Heterodera glycines</i> CN							
		<i>Meloidogyne hapla</i> RKN (4)							
		<i>Meloidogyne incognita</i> RKN (1)							

GH5 (cel)	Cellulase (EC:3.2.1.4)	<i>Aphelenchus avenae</i> (29)			In esophageal gland cells of juveniles and adults	Monophyly of PPN. Good correlation with species taxonomy except <i>A.avenae</i> cellulase more closely related to Mi-ENG2	<i>Apriona germari*</i> (37), <i>Psacothea hilaris*</i> (38), then <i>Cytophaga hutchinsonii</i> (Bacteroidetes)	Most of duplications took place after the separation of Cyst and RKN but before the further differentiation of species within these lineages.	Two intron positions are shared between CN, Anguinidae and Radopholinae, two positions are shared between CN and RKN, one position is shared between Anguinidae and Radopholinae, one position is shared between CN and Radopholinae, one position is shared between CN, RKN and Radopholinae and one position is shared by all the four lineages.
		<i>Ditylenchus africanus</i> (39)							
		<i>Ditylenchus destructor</i>							
		<i>Globodera pallida</i> CN							
		<i>Globodera rostochiensis</i> *CN (40-42)			Y				
		<i>Globodera tabacum</i> CN (43)			In subventral gland cells, during migratory stages and in pre-hatched eggs, also in parasitic stages, not in J3, J4 and adults but in late J4 males.				
		<i>Heterodera glycines</i> *CN (32, 40, 42, 44-47)		Y	In subventral gland cells, during migratory stages and in pre-hatched J2s				
		<i>Heterodera schachtii</i> CN (48, 49)		Y					
		<i>Meloidogyne arenaria</i> RKN							
		<i>Meloidogyne chitwoodii</i> RKN							
		<i>Meloidogyne hapla</i> RKN (26)							
		<i>Meloidogyne incognita</i> *RKN (1, 50-52)		Y	In subventral gland cells				
		<i>Meloidogyne javanica</i> RKN							
		<i>Pratylenchus coffeae</i> (39)							
		<i>Pratylenchus penetrans</i> (53)							
		<i>Radopholus similis</i> (54)			In pharyngeal gland cells				
GH5 (xyl)	Endo-1,4-beta-xylanase (E.C:3.2.1.8)	<i>Meloidogyne arenaria</i> RKN				Monophyly of PPN , only found in RKN and <i>Radopholus</i>	<i>Clostridium acetobutylicum</i> (Firmicutes)	Not possible to state precisely due to the absence in CN. Duplications started before separation of <i>M. hapla</i> and <i>M. incognita</i> and continued in <i>M. incognita</i>	Well conserved across different species suggesting early intron gain.
		<i>Meloidogyne chitwoodii</i> RKN							
		<i>Meloidogyne hapla</i> RKN (26)							
		<i>Meloidogyne incognita</i> *RKN (1, 55)			In subventral gland cells				
		<i>Meloidogyne javanica</i> RKN							
		<i>Radopholus similis</i> * (56, 57)	Reduced infection.		In gland cells				
Expansin-like	Loosening of plant cell wall	<i>Bursaphelenchus mucronatus</i> (1)				Monophyly of PPN, including <i>Xiphinema index</i> (Clade I PPN)	Actinomycetales: <i>Amycolatopsis mediterranei</i> , <i>Actinosynnema mirum</i> , <i>Streptomyces lavendulae</i>	Duplications started before separation of RKN and CN and most occurred independently in both lineages after their separation.	
		<i>Bursaphelenchus xylophilus</i> (1)							
		<i>Ditylenchus africanus</i> (58)			In gland cells area				
		<i>Globodera rostochiensis</i> * CN (59, 60)		Y	In subventral esophageal gland cells				
		<i>Meloidogyne hapla</i> RKN							
		<i>Meloidogyne incognita</i> RKN (35)							
		<i>Meloidogyne javanica</i> RKN							
		<i>Xiphinema index</i>							

* Activity experimentally shown | CN: cyst nematode | RKN: root-knot nematode

Figure S1



Meloidogyne incognita, *Meloidogyne hapla*, *Globodera pallida*, *Aphelenchus avenae*, *Bursaphelenchus xylophilus*, *Zeldia punctata*, *Panagrellus redivivus*, *Strongyloides ratti*, *Steinernema carpocapsae*, *Caenorhabditis elegans*, *Caenorhabditis briggsae*, *Haemonchus contortus*, *Pristionchus pacificus*, *Brugia malayi*, *Ascaris suum*, *Rhigonema thysanophora*, *Oxyuris equi*, *Plectus aquatilis*, *Monhystera riemannii*, *Chromadorita cf. leuckarti*, *Xiphinema index*, *Mononchus truncatus*, *Mermis nigrescens*, *Trichinella spiralis*, *Trichodorus cylindricus*, *Enoplus meridionalis*

Fig. S2A

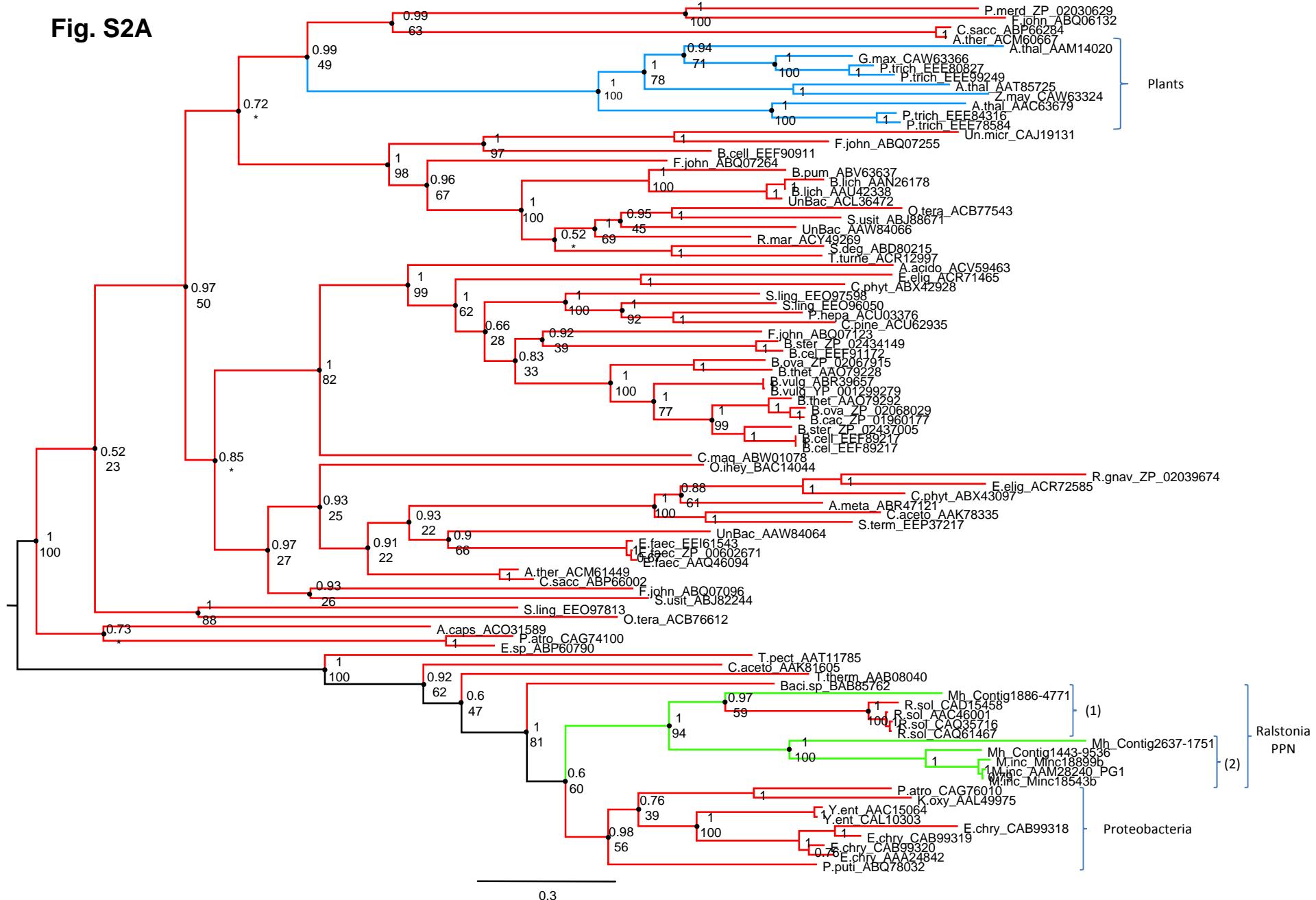


Fig. S2B

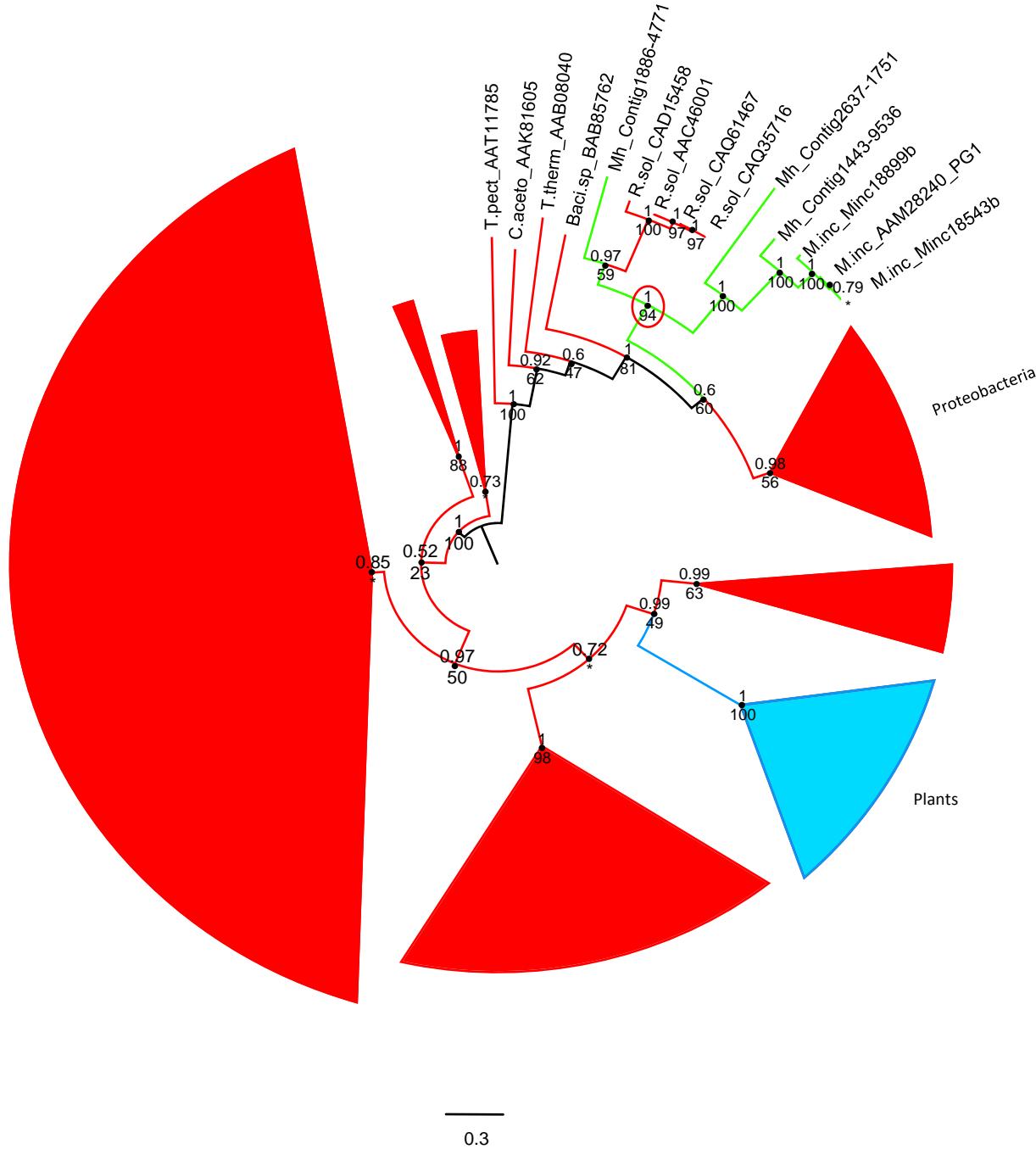


Fig. S3A

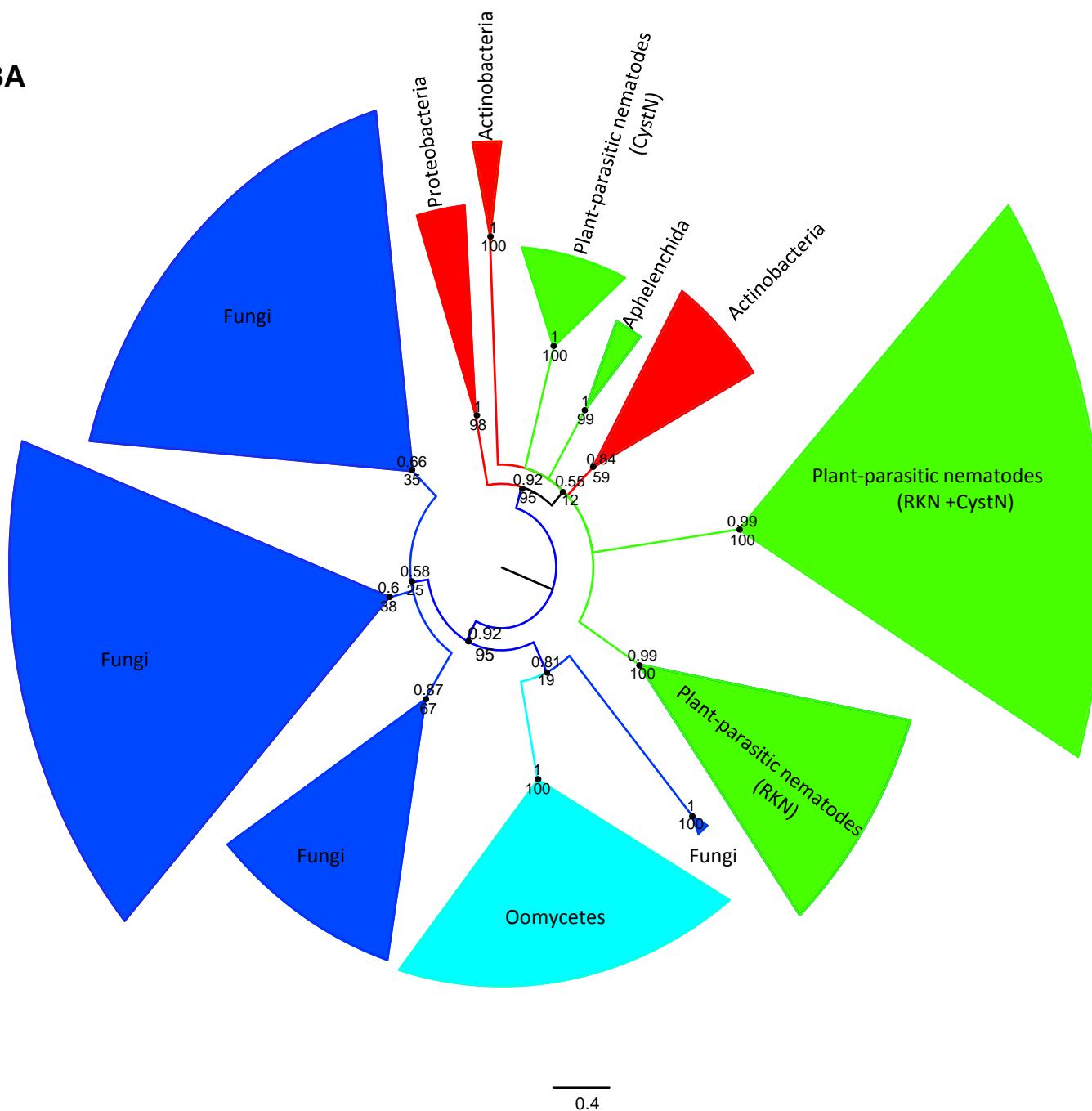


Fig. S3B

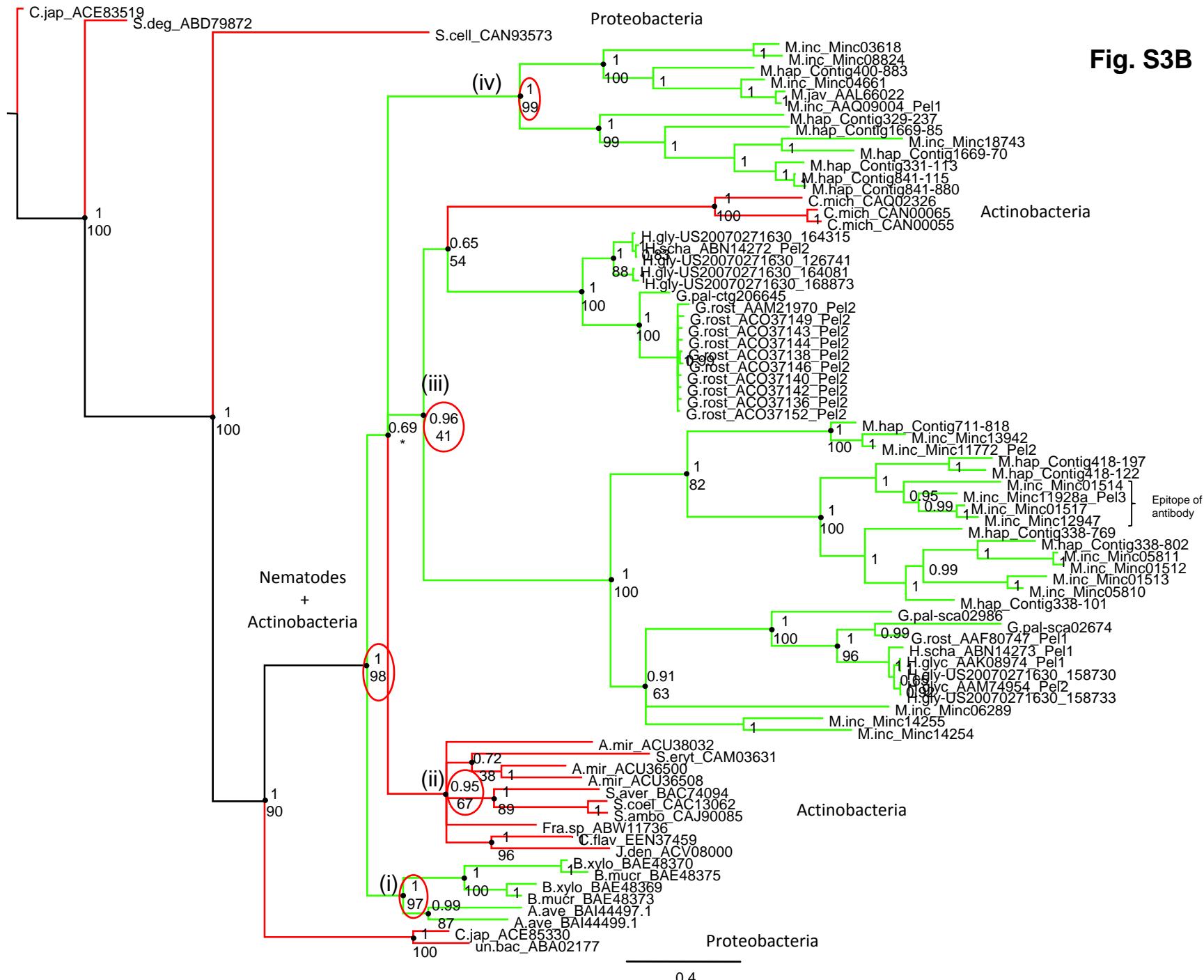


Fig. S3C

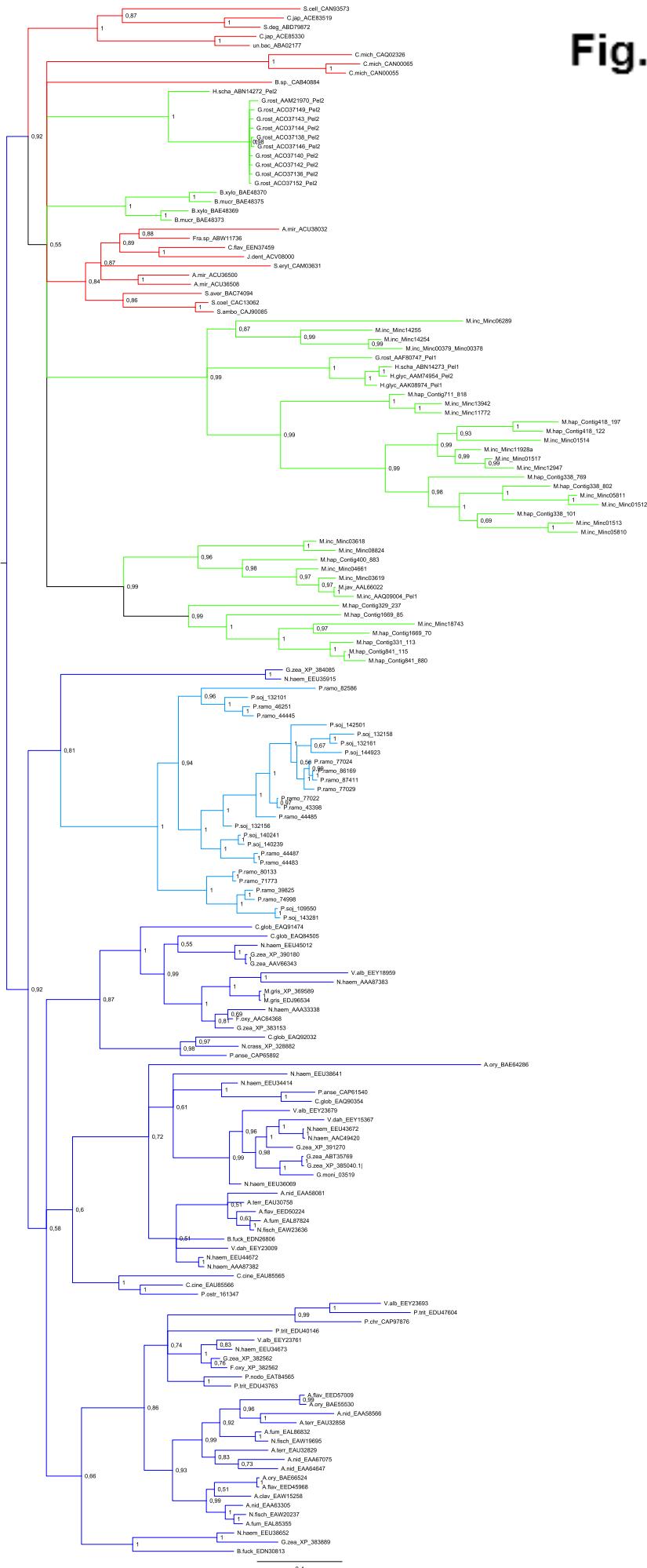


Fig. S4

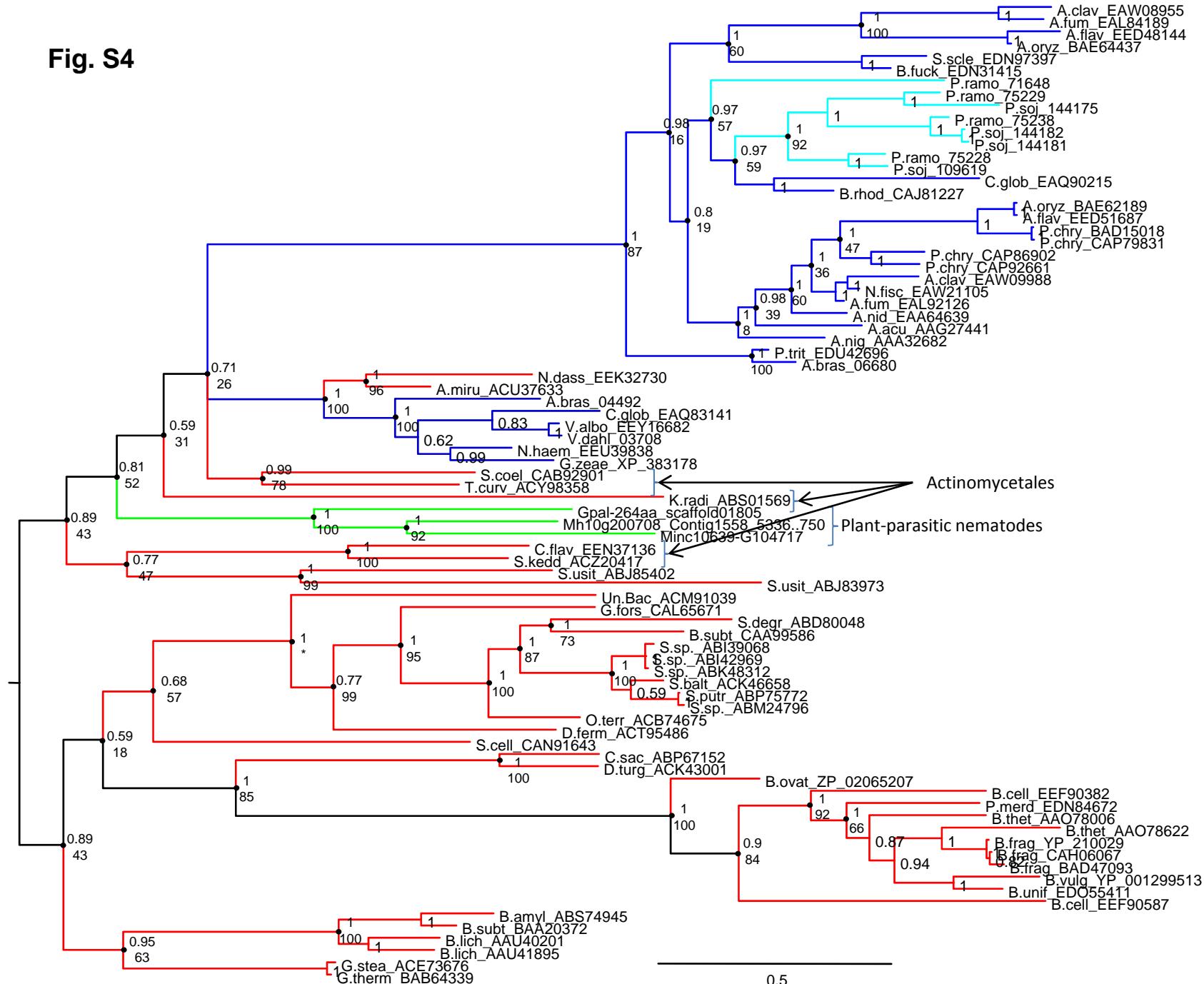


Fig. S5A

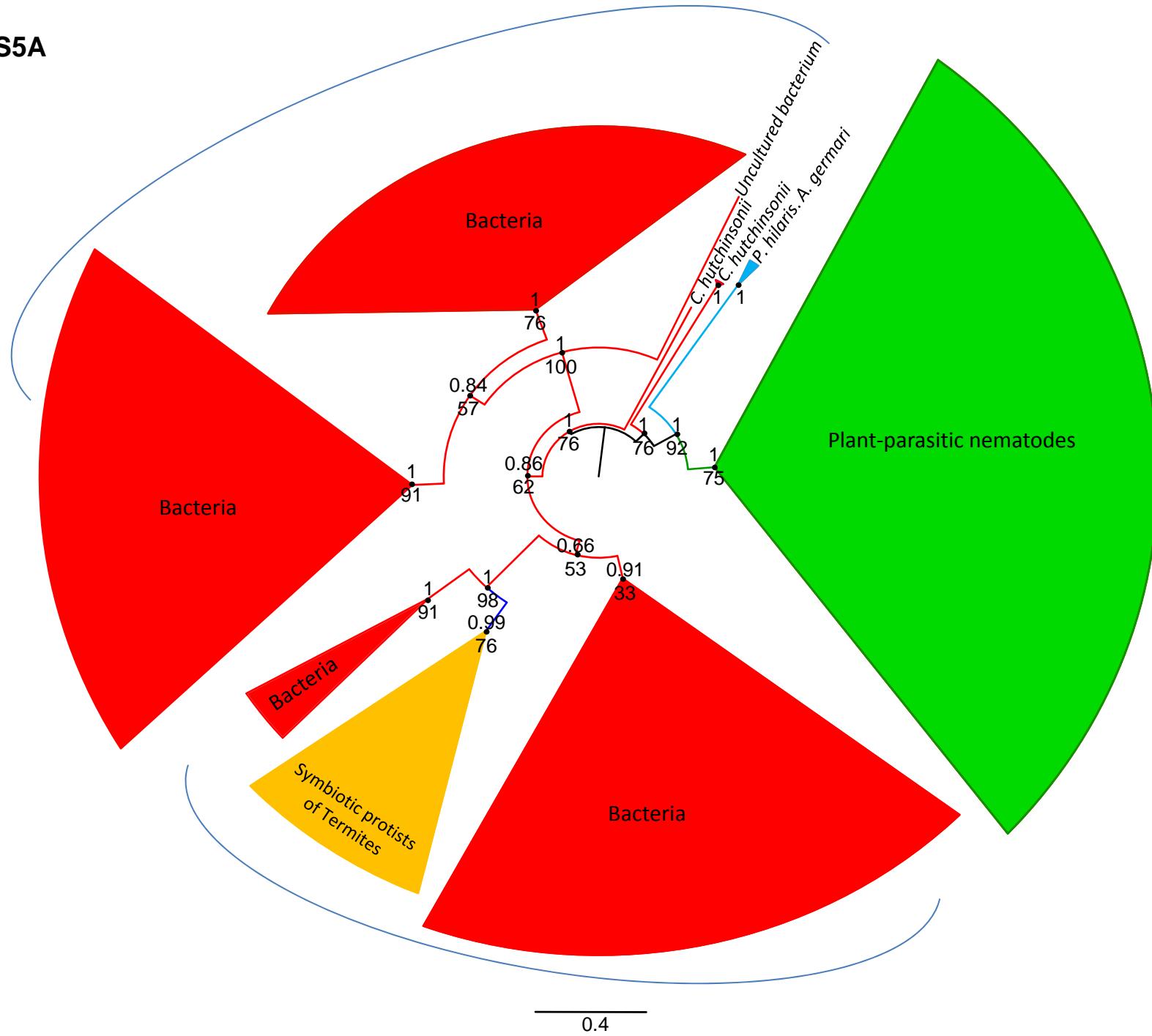


Fig. S5B

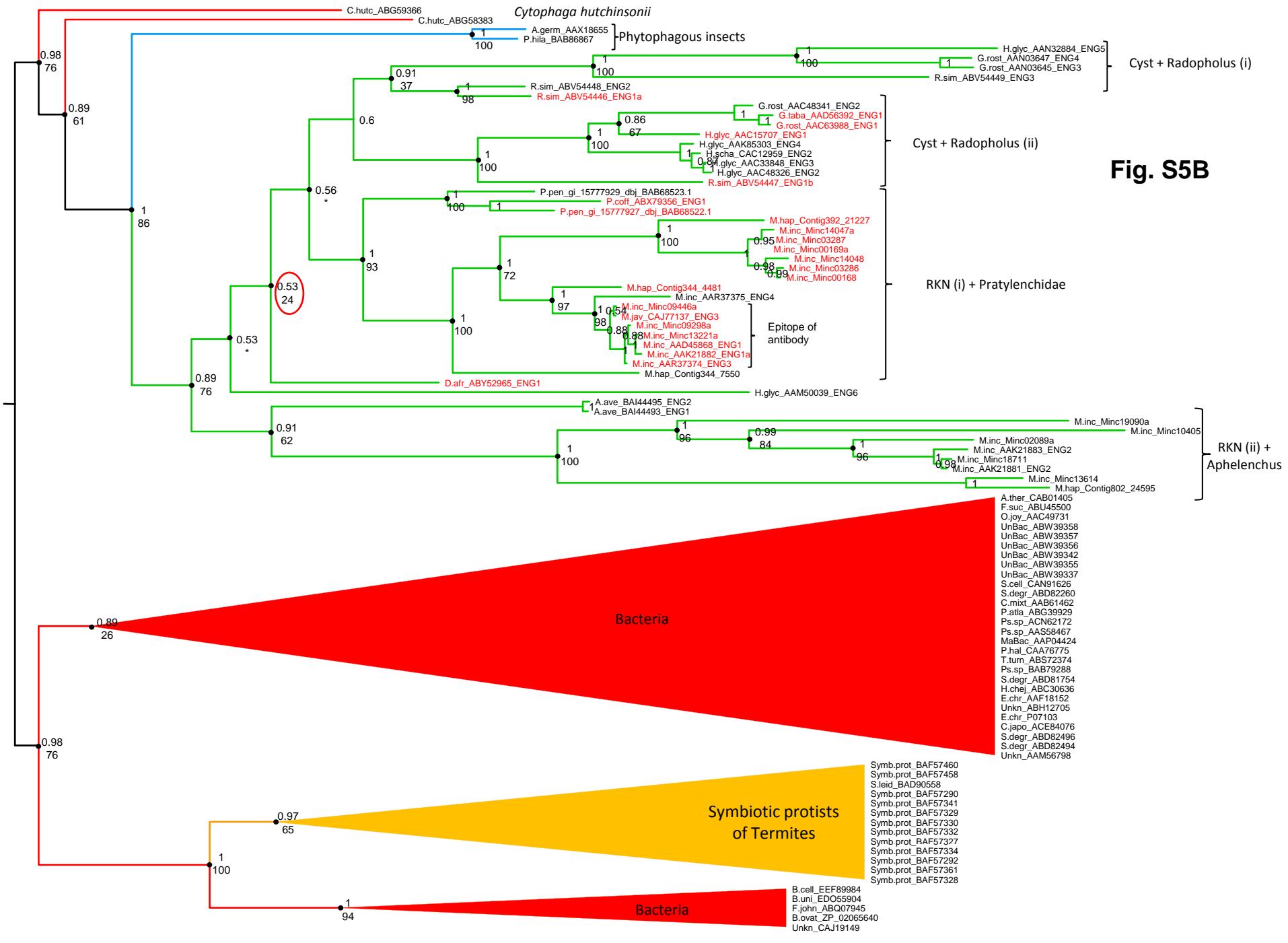


Fig. S5C

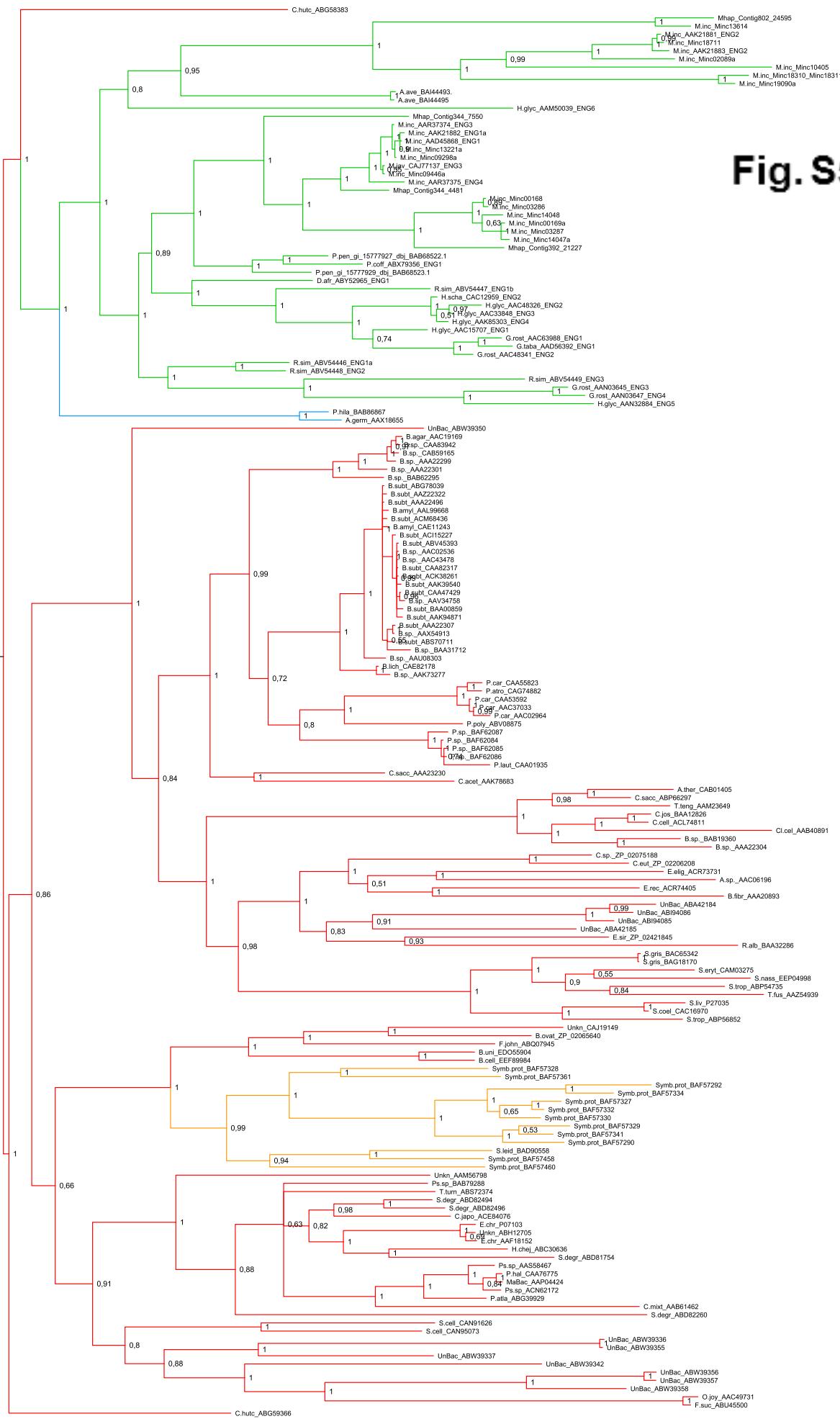


Fig. S6A

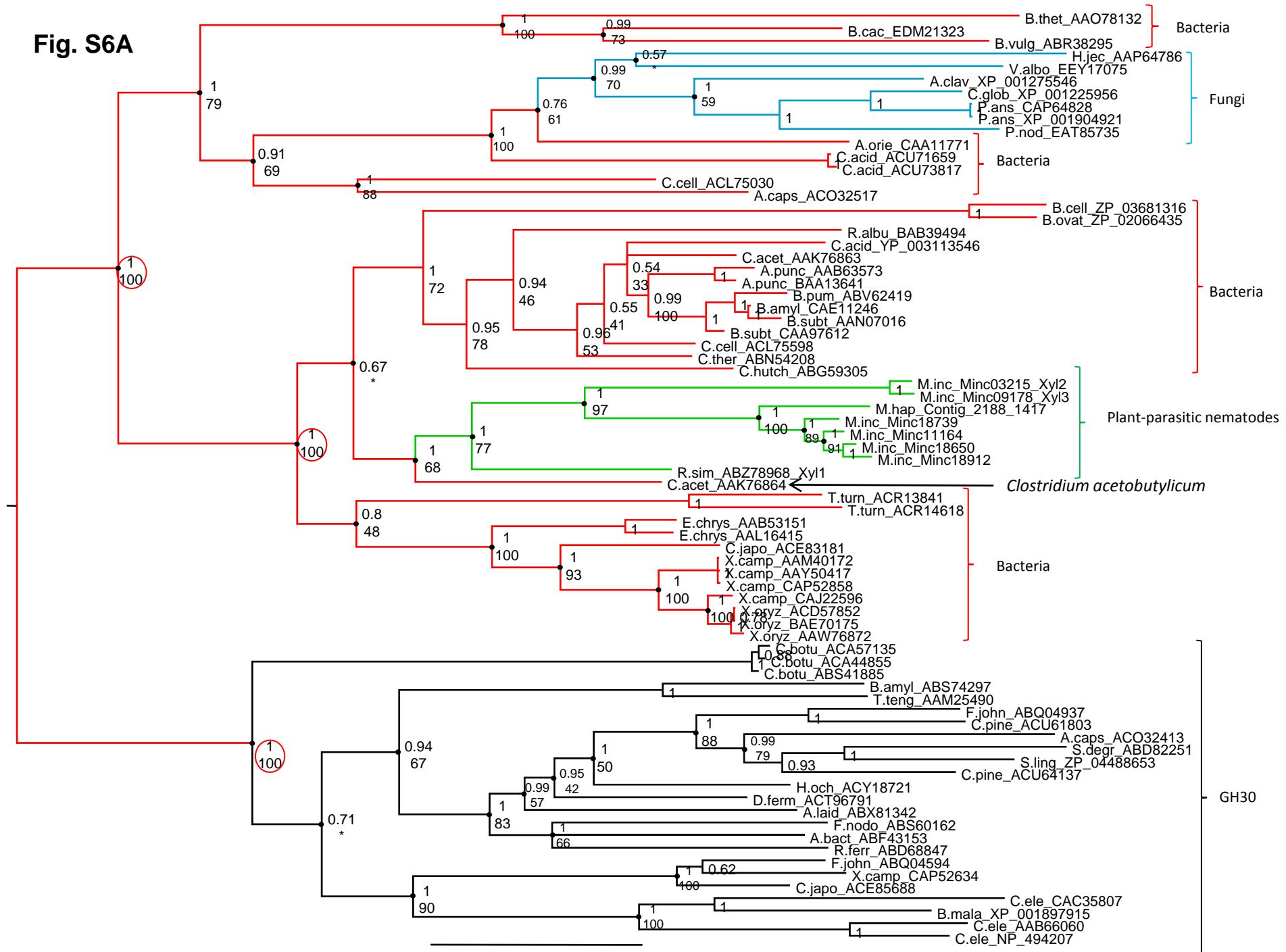


Fig. S6B

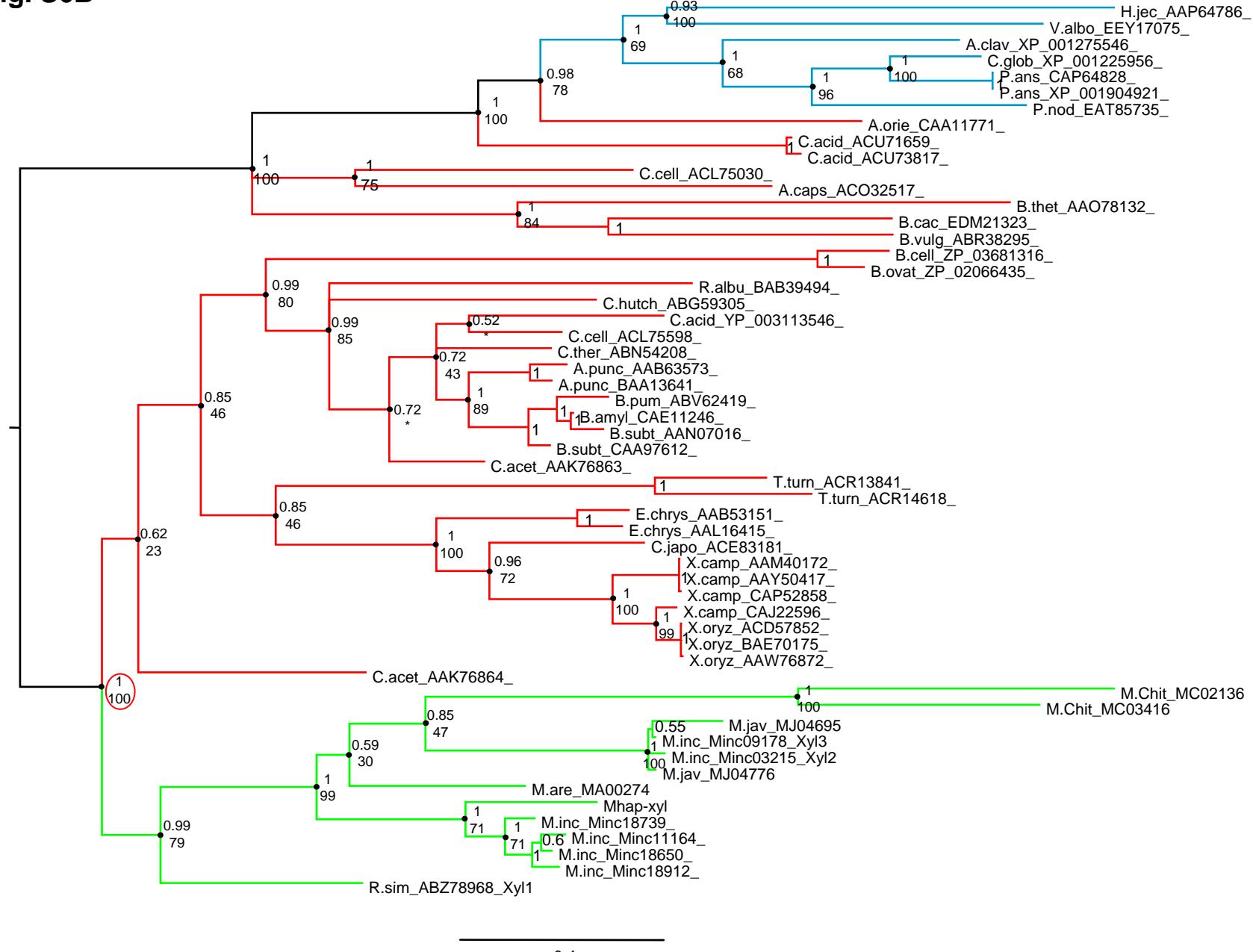


Fig. S7A

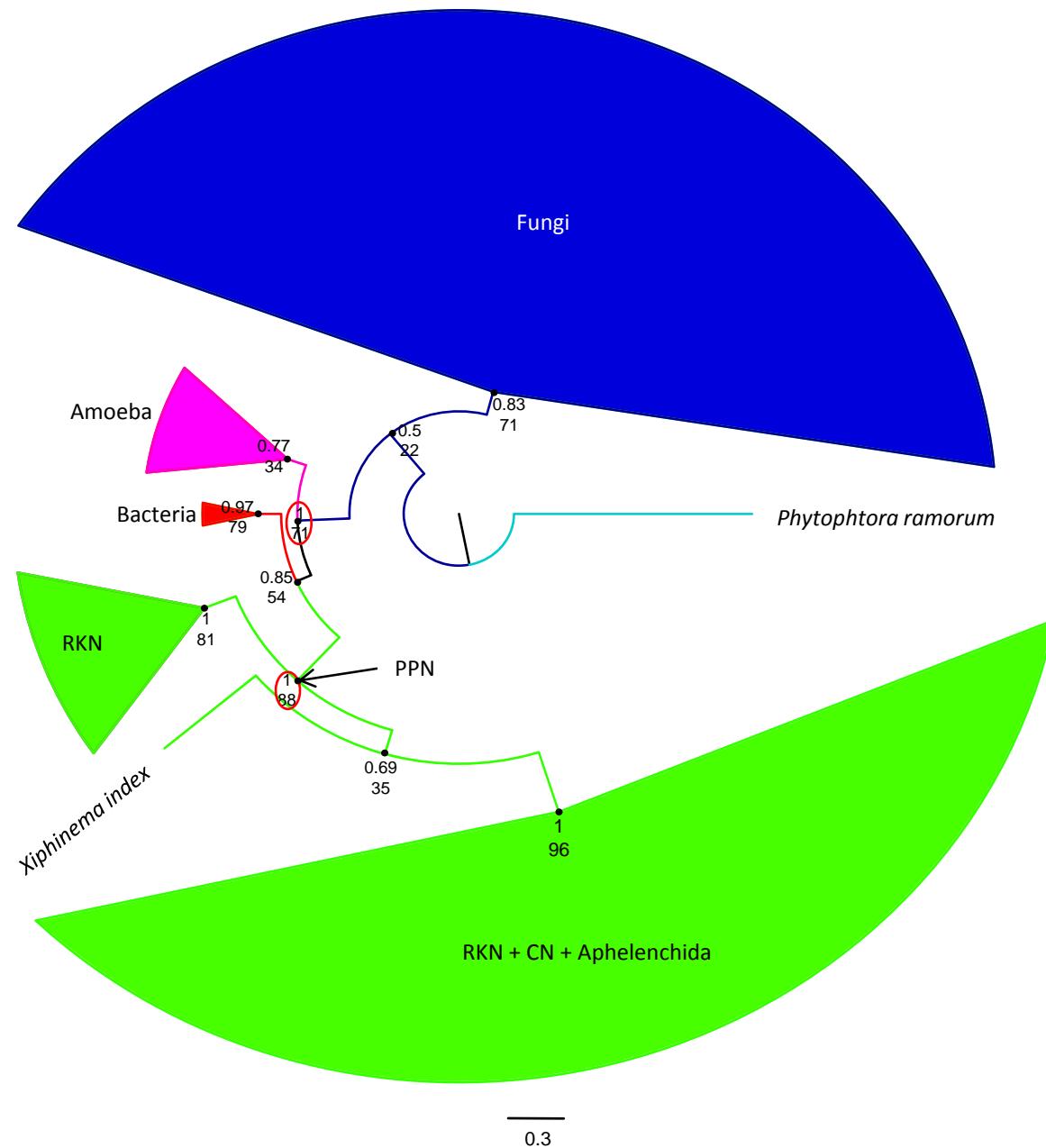


Fig. S7B

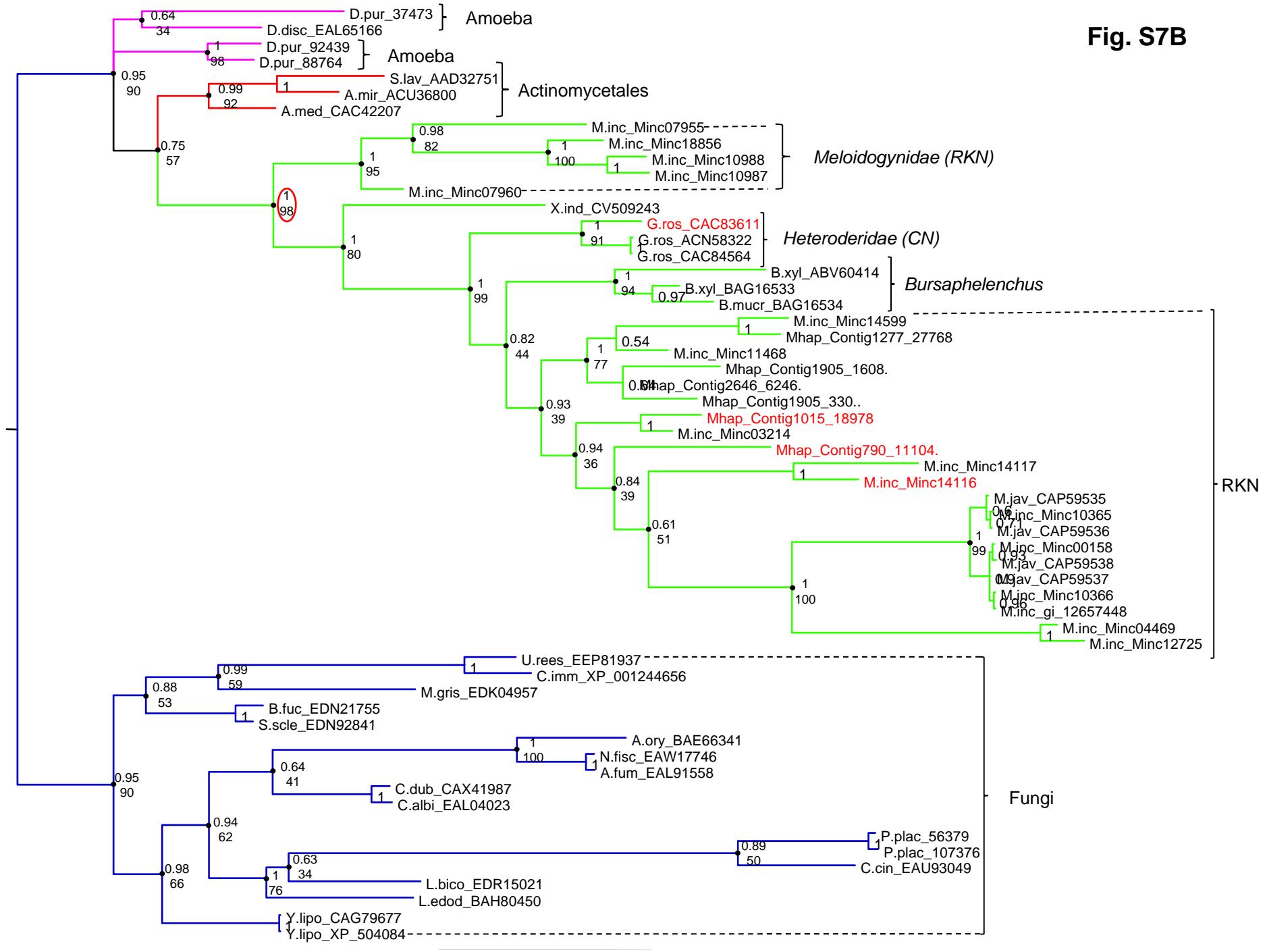


Fig. S8A

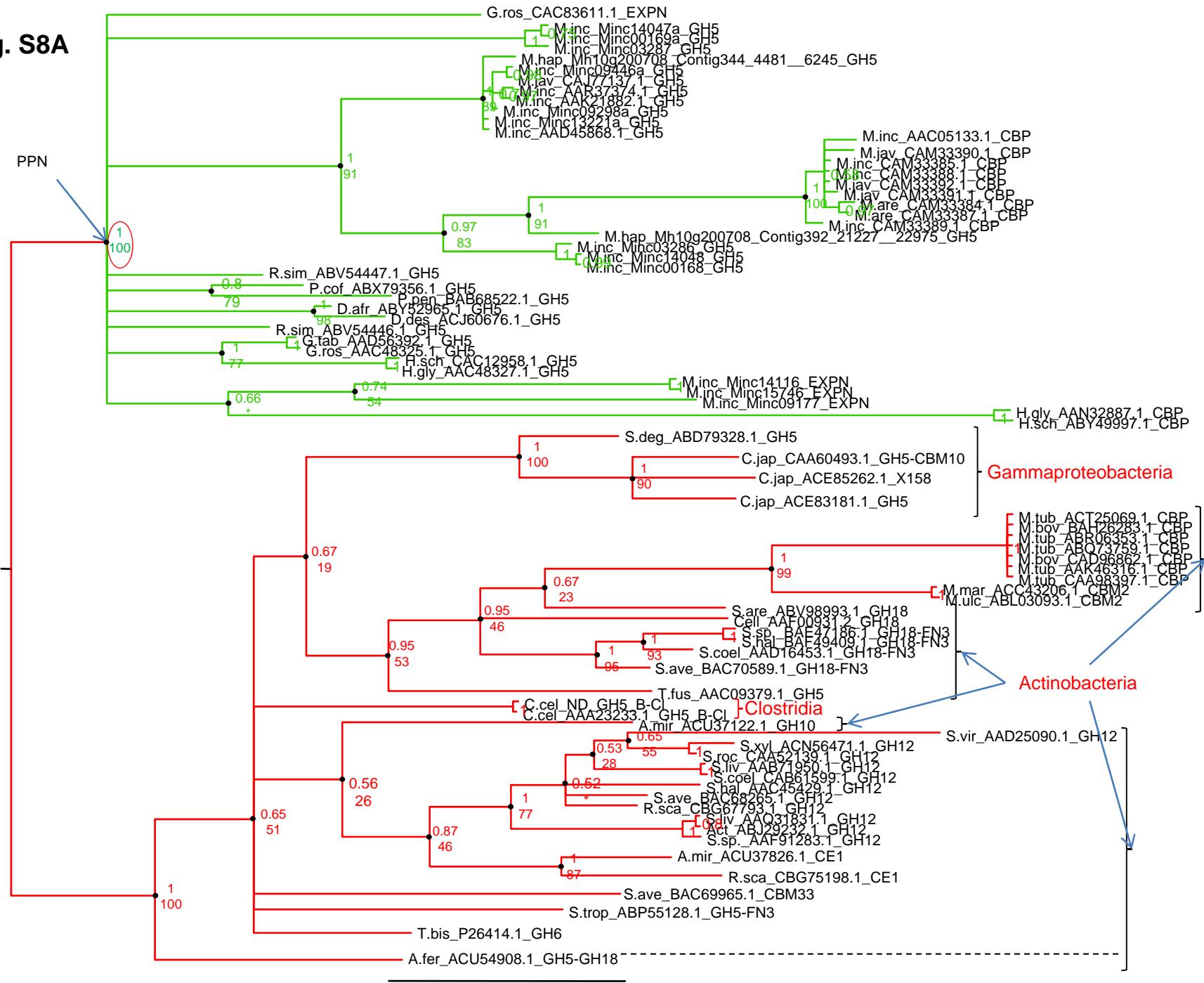
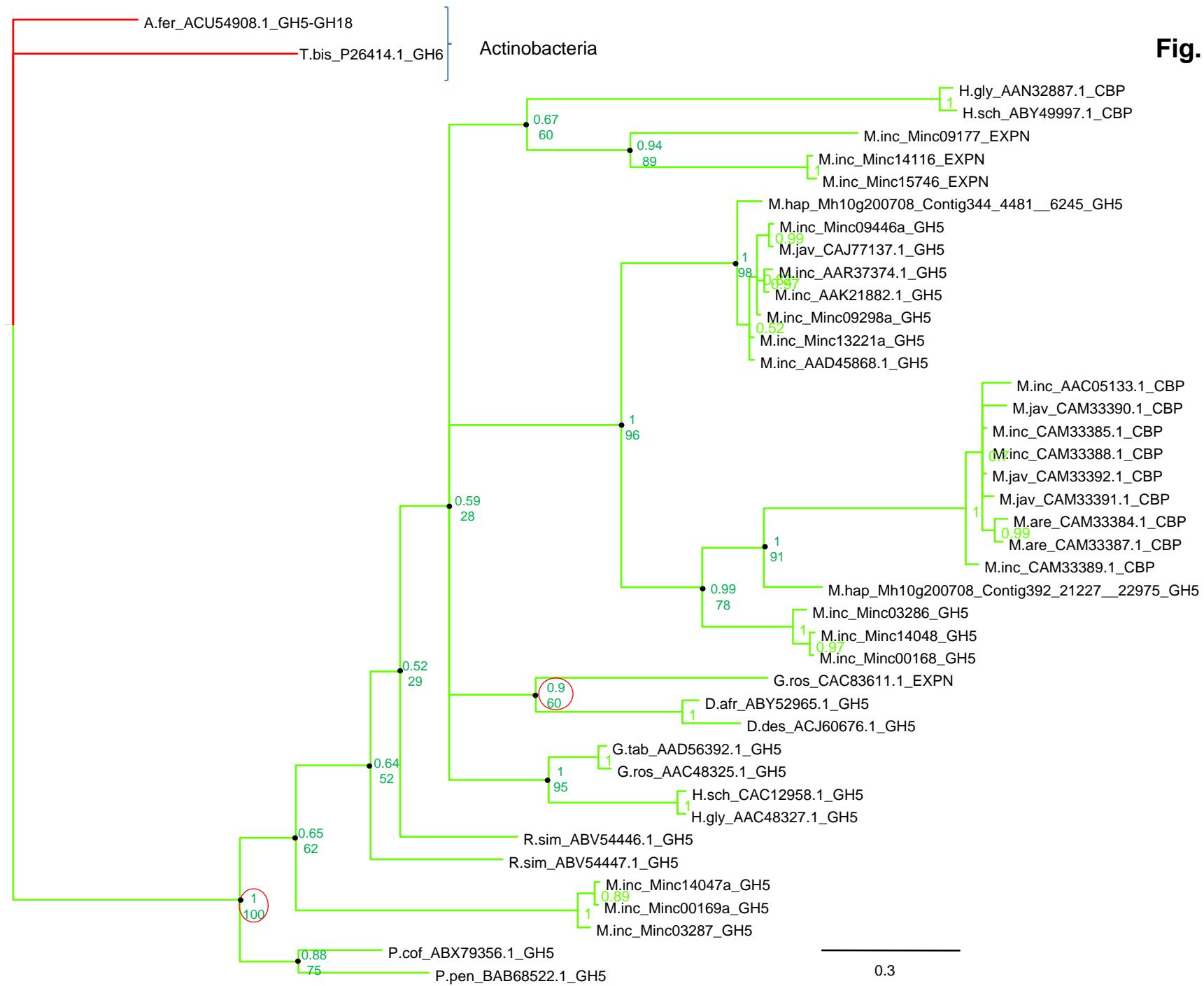


Fig. S8B



Tables S2A-S2F. Statistics of comparison of constrained and non-constrained tree topologies.

Table S2A. GH28 polygalacturonases.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-39.3	0.980	0.779	0.772	1.000	0.905	1.000	0.905	1.000
2	11	39.3	0.227	0.069	0.072	8,0E-18	0.095	0.738	0.095	0.559
3	3	47.7	0.082	0.017	0.017	2,0E-21	0.036	0.643	0.036	0.303
4	6	54.0	0.107	0.030	0.028	4,0E-24	0.045	0.595	0.045	0.330
5	5	59.6	0.080	0.031	0.033	1,0E-26	0.046	0.546	0.046	0.328
6	4	61.6	0.087	0.020	0.021	2,0E-27	0.033	0.497	0.033	0.249
7	7	62.1	0.077	0.028	0.031	1,0E-27	0.036	0.500	0.036	0.267
8	10	63.7	0.049	0.009	0.010	2,0E-28	0.019	0.477	0.019	0.177
9	9	64.8	0.053	0.011	0.010	8,0E-29	0.021	0.468	0.021	0.180
10	8	76.8	0.029	0.005	0.005	4,0E-34	0.011	0.370	0.011	0.101
11	2	97.4	0.005	0.001	0.001	5,0E-43	0.003	0.247	0.003	0.023
12	12*	207.2	9,0E-69	1,0E-20	0	1,0E-90	0	0,017	0	0
13	22*	246.9	3,0E-05	4,0E-06	0	6,0E-108	0	0.004	0	0
14	17*	249.5	7,0E-06	3,0E-07	0	4,0E-109	0	0.005	0	0
15	21*	259.7	0.001	4,0E-07	0	2,0E-113	0	0.003	0	0
16	19*	276.8	0.004	4,0E-05	0	6,0E-121	0	0.002	0	0
17	15*	285.6	0.001	2,0E-06	0	1,0E-124	0	0.001	0	0
18	20*	287.0	2,0E-08	2,0E-07	0	2,0E-125	0	0.002	0	0
19	18*	294.2	2,0E-04	3,0E-05	0	2,0E-128	0	0.001	0	0
20	16*	298.0	1,0E-04	6,0E-05	0	4,0E-130	0	4,0E-04	0	0
21	14*	303.6	8,0E-39	1,0E-14	0	1,0E-132	0	0.001	0	0
22	13*	1987.5	4,0E-05	1,0E-05	0	0	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode and plant GH28 homologs in a monophyletic group).

Table S2B. PL3 pectate lyases.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-10.5	0.750	0.667	0.677	1.000	0.678	0.996	0.678	0.998
2	12*	10.5	0.367	0.322	0.318	3,0E-05	0.322	0.965	0.322	0.927
3	15*	154.4	0.001	5,0E-04	0.001	9,0E-68	0.001	0.139	0.001	0.015
4	21*	169.3	0.003	0.001	4,0E-04	3,0E-74	0.001	0.112	0.001	0.013
5	10	171.2	0.004	4,0E-04	2,0E-04	4,0E-75	0.001	0.095	0.001	0.010
6	19*	172.6	0.001	4,0E-04	4,0E-04	1,0E-75	4,0E-04	0.088	4,0E-04	0.006
7	11	176.5	0.003	0.001	5,0E-04	2,0E-77	0.002	0.091	0.002	0.016
8	9	184.0	0.002	4,0E-04	0.001	1,0E-80	0.001	0.076	0.001	0.012
9	2	189.7	0.001	2,0E-04	2,0E-04	4,0E-83	2,0E-04	0.046	2,0E-04	0.004
10	22*	192.2	0.006	0.001	0.001	3,0E-84	0.001	0.054	0.001	0.008
11	16*	193.6	0.001	3,0E-04	1,0E-04	8,0E-85	1,0E-04	0.052	1,0E-04	0.001
12	6	194.7	0.001	3,0E-04	3,0E-04	3,0E-85	0.001	0.044	0.001	0.008
13	4	201.2	3,0E-04	4,0E-05	0	4,0E-88	0	0.041	0	0.002
14	7	202.3	3,0E-04	1,0E-04	2,0E-04	1,0E-88	2,0E-04	0.035	2,0E-04	0.002
15	5	202.7	0.001	5,0E-04	3,0E-04	9,0E-89	2,0E-04	0.043	2,0E-04	0.001
16	3	204.6	0.001	2,0E-04	1,0E-04	1,0E-89	3,0E-04	0.038	3,0E-04	0.006
17	14*	204.9	0.002	2,0E-04	0	1,0E-89	2,0E-04	0.038	2,0E-04	0.003
18	17*	215.1	2,0E-05	2,0E-05	0	4,0E-94	4,0E-04	0.022	4,0E-04	0.003
19	8	218.6	2,0E-04	5,0E-05	0	1,0E-95	3,0E-05	0.023	3,0E-05	0.001
20	20*	219.5	2,0E-05	3,0E-05	1,0E-04	5,0E-96	2,0E-04	0.034	2,0E-04	0.004
21	18*	243.4	0.003	2,0E-04	3,0E-05	2,0E-106	0	0.010	0	4,0E-04
22	13*	1317.9	3,0E-04	1,0E-05	0	0	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode, Oomycete and Fungal PL3 homologs in a monophyletic group).

Table S2C. GH43 candidate arabinanases.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-43.1	0.995	0.892	0.892	1.000	0.955	1.000	0.955	1.000
2	5	43.1	0.088	0.035	0.034	2,0E-19	0.045	0.604	0.045	0.350
3	6	48.9	0.029	0.014	0.014	6,0E-22	0.022	0.501	0.022	0.168
4	12*	51.6	0.036	0.013	0.014	4,0E-23	0.019	0.469	0.019	0.190
5	8	64.3	0.036	0.011	0.012	1,0E-28	0.016	0.327	0.016	0.129
6	3	68.4	0.059	0.018	0.015	2,0E-30	0.023	0.285	0.023	0.168
7	4	70.8	0.006	0.002	0.001	2,0E-31	0.004	0.256	0.004	0.033
8	11	73.4	0.031	0.006	0.007	1,0E-32	0.011	0.224	0.011	0.096
9	7	78.1	0.005	0.001	0.001	1,0E-34	0.002	0.187	0.002	0.022
10	9	83.5	0.020	0.004	0.005	5,0E-37	0.007	0.176	0.007	0.060
11	22*	88.2	0.007	0.001	5,0E-04	5,0E-39	0.002	0.089	0.002	0.031
12	2	91.9	0.002	0.001	0.001	1,0E-40	0.001	0.108	0.001	0.017
13	18*	100.4	0.004	0.001	4,0E-04	3,0E-44	0.003	0.059	0.003	0.034
14	17*	106.0	4,0E-04	7,0E-05	4,0E-05	9,0E-47	0.001	0.046	0.001	0.007
15	14*	107.3	0.008	4,0E-04	2,0E-04	3,0E-47	0.001	0.038	0.001	0.018
16	10	121.4	0.007	0.001	0.001	2,0E-53	0.003	0.034	0.003	0.027
17	19*	124.3	0.002	3,0E-04	3,0E-04	1,0E-54	0.001	0.022	0.001	0.008
18	20*	127.7	0.004	2,0E-04	1,0E-04	3,0E-56	0.001	0.013	0.001	0.008
19	15*	133.0	0.001	1,0E-04	1,0E-04	2,0E-58	3,0E-04	0.014	3,0E-04	0.005
20	21*	134.8	0.002	3,0E-04	0.001	3,0E-59	0.001	0.014	0.001	0.011
21	16*	155.6	0.001	3,0E-05	0	3,0E-68	0	0.001	0	1,0E-04
22	13*	248.9	2,0E-46	3,0E-16	0	8,0E-109	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode, Oomycete and Fungal GH43 homologs in a monophyletic group).

Table S2D. GH5 cellulases.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-43.9	0.978	0.832	0.828	1.000	0.912	1.000	0.912	1.000
2	7	43.9	0.156	0.069	0.070	9,0E-20	0.088	0.636	0.088	0.485
3	11	53.6	0.124	0.045	0.048	5,0E-24	0.062	0.525	0.062	0.399
4	9	54.2	0.057	0.013	0.013	3,0E-24	0.023	0.521	0.023	0.203
5	4	59.7	0.027	0.007	0.007	1,0E-26	0.015	0.457	0.015	0.127
6	6	66.4	0.028	0.007	0.007	2,0E-29	0.013	0.380	0.013	0.122
7	8	68.1	0.051	0.012	0.013	3,0E-30	0.022	0.368	0.022	0.173
8	5	77.2	0.020	0.005	0.004	3,0E-34	0.009	0.275	0.009	0.094
9	10	80.6	0.041	0.007	0.008	1,0E-35	0.016	0.246	0.016	0.153
10	3	108.8	0.005	0.001	0.001	5,0E-48	0.004	0.121	0.004	0.025
11	2	145.0	9,0E-05	3,0E-05	0	1,0E-63	3,0E-04	0.014	3,0E-04	0.004
12	12*	170.7	7,0E-59	7,0E-19	0	7,0E-75	0	0.002	0	0
13	20*	217.6	1,0E-50	5,0E-17	0	3,0E-95	0	6,0E-05	0	0
14	15*	218.3	3,0E-09	3,0E-07	0	2,0E-95	0	0	0	0
15	18*	222.1	2,0E-05	2,0E-06	0	3,0E-97	0	4,0E-05	0	0
16	21*	232.8	2,0E-80	2,0E-22	0	8,0E-102	0	0	0	0
17	14*	242.9	3,0E-30	3,0E-14	0	3,0E-106	0	0	0	0
18	17*	244.6	4,0E-80	2,0E-21	0	6,0E-107	0	0	0	0
19	19*	251.5	3,0E-148	2,0E-34	0	6,0E-110	0	0	0	5,0E-05
20	16*	260.2	5,0E-42	3,0E-15	0	1,0E-113	0	0	0	0
21	22*	312.9	4,0E-72	2,0E-20	0	1,0E-136	0	0	0	0
22	13*	603.6	2,0E-07	7,0E-07	0	8,0E-263	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode, Insect and symbiotic protists GH5 cellulase homologs in a monophyletic group).

Table S2E. GH5 xylanases.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-29.0	0.970	0.677	0.669	1.000	0.891	1.000	0.891	1.000
2	8	29.0	0.222	0.075	0.077	3,0E-13	0.109	0.759	0.109	0.556
3	11	30.2	0.162	0.049	0.053	8,0E-14	0.071	0.740	0.071	0.451
4	3	33.5	0.220	0.055	0.057	3,0E-15	0.089	0.697	0.089	0.526
5	4	36.7	0.203	0.044	0.041	1,0E-16	0.074	0.636	0.074	0.440
6	2	43.5	0.070	0.010	0.009	1,0E-19	0.027	0.556	0.027	0.228
7	9	45.1	0.077	0.022	0.022	2,0E-20	0.042	0.532	0.042	0.266
8	6	47.9	0.065	0.009	0.008	2,0E-21	0.021	0.496	0.021	0.175
9	5	50.0	0.115	0.025	0.025	2,0E-22	0.060	0.479	0.060	0.333
10	10	50.0	0.099	0.025	0.026	2,0E-22	0.042	0.470	0.042	0.237
11	7	60.5	0.059	0.012	0.014	5,0E-27	0.025	0.356	0.025	0.173
12	12*	287.9	4,0E-53	3,0E-17	0	9,0E-126	0	0	0	0
13	14*	299.9	4,0E-05	2,0E-06	0	6,0E-131	0	0	0	0
14	17*	304.1	9,0E-06	7,0E-06	0	9,0E-133	0	0	0	0
15	22*	323.2	3,0E-79	3,0E-21	0	4,0E-141	0	0	0	0
16	20*	329.9	2,0E-05	3,0E-06	0	5,0E-144	0	0	0	0
17	21*	330.4	2,0E-07	6,0E-07	0	3,0E-144	0	0	0	0
18	19*	333.7	2,0E-36	4,0E-15	0	1,0E-145	0	0	0	0
19	16*	336.5	1,0E-04	3,0E-06	0	7,0E-147	0	0	0	0
20	18*	340.6	2,0E-06	4,0E-07	0	1,0E-148	0	0	0	0
21	15*	348.0	0.006	2,0E-06	0	7,0E-152	0	0	0	0
22	13*	653.6	9,0E-06	1,0E-06	0	1,0E-284	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode, and fungal GH5 xylanases homologs in a monophyletic group).

Table S2F. Expansin-like proteins.

Rank	Tree	obs	au	np	bp	pp	kh	sh	wkh	wsh
1	1	-0.4	0.818	0.431	0.436	0.589	0.526	0.985	0.526	0.989
2	12*	0.4	0.733	0.353	0.357	0.411	0.474	0.986	0.474	0.981
3	7	30.5	0.205	0.048	0.049	3,0E-14	0.090	0.577	0.090	0.478
4	22*	33.3	0.199	0.040	0.037	2,0E-15	0.089	0.567	0.089	0.502
5	11	33.8	0.143	0.018	0.015	1,0E-15	0.074	0.559	0.074	0.459
6	4	34.0	0.151	0.026	0.027	1,0E-15	0.066	0.537	0.066	0.433
7	18*	34.6	0.143	0.023	0.025	6,0E-16	0.074	0.522	0.074	0.433
8	15*	40.8	0.062	0.008	0.008	1,0E-18	0.030	0.425	0.031	0.246
9	3	41.6	0.110	0.020	0.019	5,0E-19	0.045	0.402	0.045	0.301
10	2	45.1	0.041	0.010	0.010	2,0E-20	0.036	0.363	0.036	0.228
11	10	45.5	0.030	0.005	0.005	1,0E-20	0.018	0.351	0.016	0.113
12	20*	45.6	0.034	0.005	0.004	9,0E-21	0.018	0.345	0.016	0.159
13	5	56.0	0.036	0.005	0.005	3,0E-25	0.016	0.237	0.016	0.113
14	14*	56.7	0.006	0.001	0.001	1,0E-25	0.005	0.206	0.005	0.038
15	21*	56.8	0.008	5,0E-04	4,0E-04	1,0E-25	0.005	0.207	0.004	0.032
16	9	58.5	0.019	0.001	0.002	2,0E-26	0.011	0.201	0.011	0.097
17	8	70.5	0.004	4,0E-04	5,0E-04	1,0E-31	4,0E-04	0.087	4,0E-04	0.007
18	16*	76.4	0.004	5,0E-04	5,0E-04	4,0E-34	0.002	0.074	0.002	0.025
19	19*	85.0	0.001	1,0E-04	1,0E-04	7,0E-38	0.001	0.053	0.001	0.006
20	6	95.9	0.001	1,0E-04	1,0E-04	1,0E-42	0.001	0.027	0.001	0.009
21	17*	97.0	4,0E-04	6,0E-05	9,0E-06	4,0E-43	0.001	0.023	4,0E-04	0.005
22	13*	431.5	1,0E-52	2,0E-17	0	2,0E-188	0	0	0	0

* constrained trees (the constraint consisted in grouping Nematode, Amoeba and fungal EXPN candidate expansin homologs in a monophyletic group).

Fig. S9

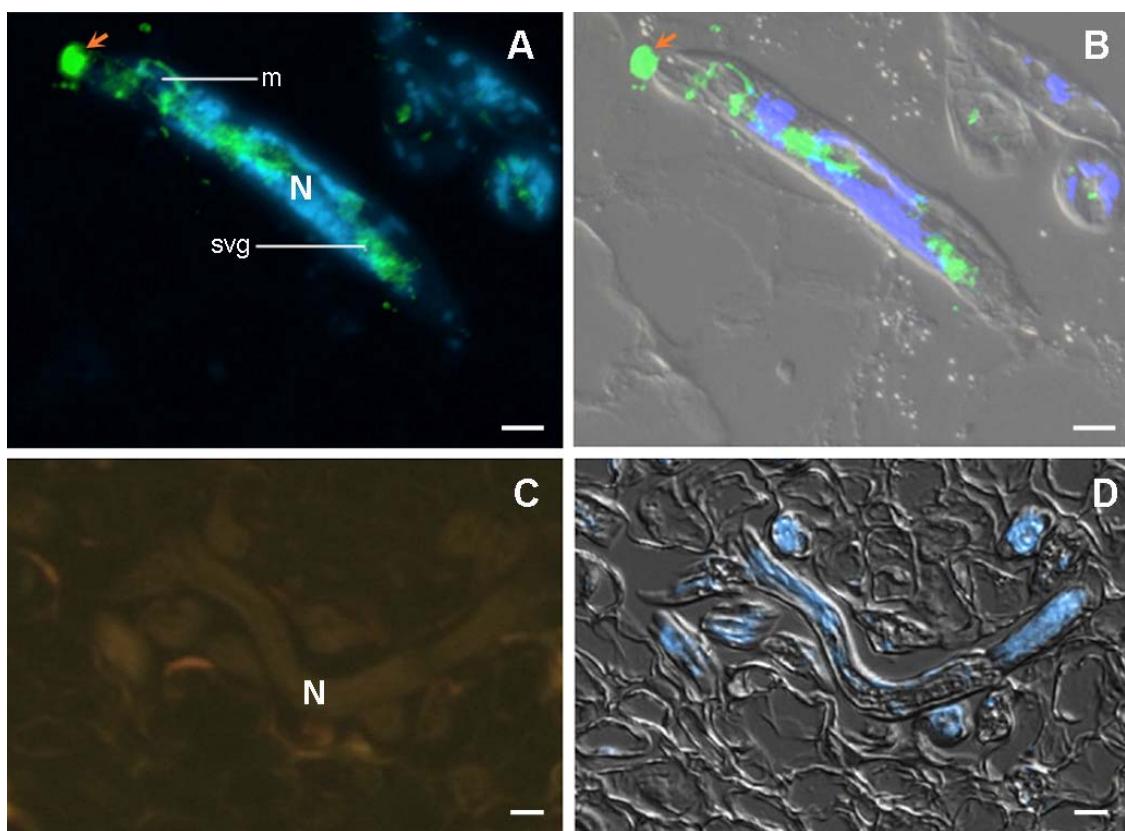
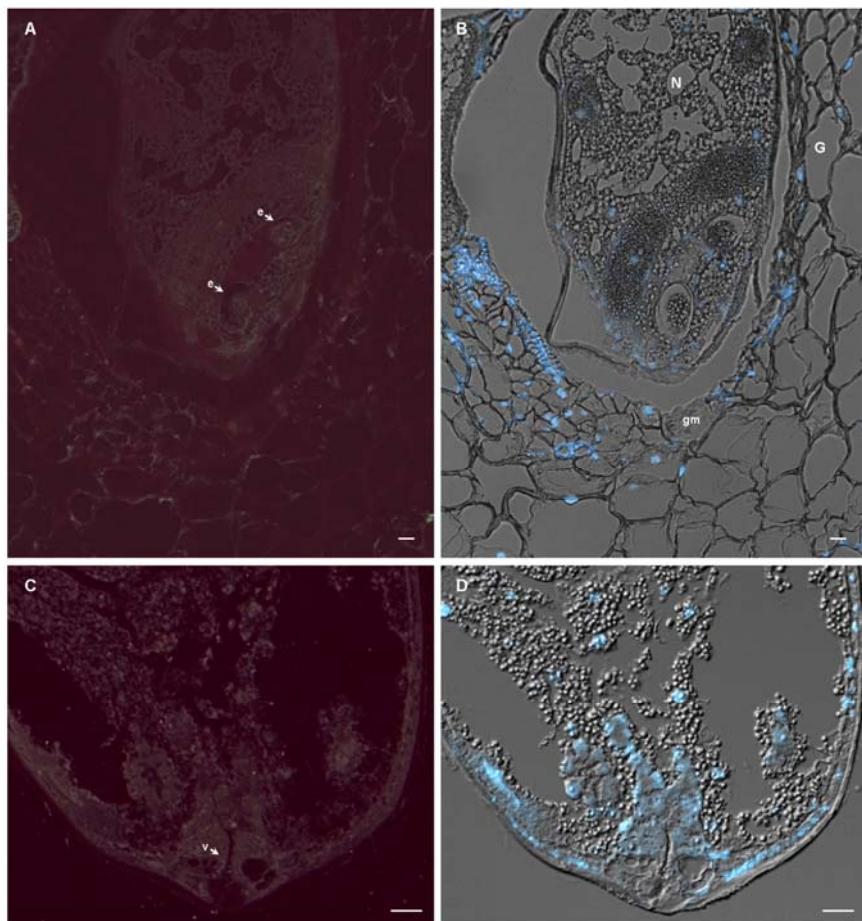


Fig S10.



A) GH28

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

10 20 30 40 50

MAFYSLF - - - LLVHLFIYS - F - SNKXXVITPQCGYQNSLPFHQSSDSFES
MKLLAIL - - - QIILIKLILLE - VAFSQXXTISPPCSCLNVKPNFGTN - - - S
MKYYILSAILKRIILIKIIFLETTNFALSQXXTITPPCSCSNVKPNFGTN - - - S
MKYYILSAILKRIILIKIIFLETTNFALSQXXTITPPCSCSNVKPNFGTN - - - S

Fig. S11

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

60 70 80 90 100

NIPLNLCVPTLAYDQTTTVTLVWNKPNNYSNIDYNIYMNGEL - IGGAKENNFKN
NIPQQLCVPSLAYDQTTSIWLTWNKPDNYENIVDFNIYMNGKXXKIGNSKTNAAIN
NIPQQLCVPPPLAYDQKSVWLWNKPDNYENIADFNVY MAGKXXKIGSAKANSAVN
NIPQQLCVPPPLAYDQKSVWLWNKPDNYENIADFNVY MAGKXXKIGSAKANSAVN

Mh10g200708_Contig1886:4771..7593 SPAFPYINLFYNEDKFNFHTTKITYHSFLVKNLEEPNKIYSFTVRSVDEENGNESKDS
Mh10g200708_Contig1443:9536..11966 TL SG PY IQNF YKNDL NN FHT K I L F T T Y L V K G L N P N T I Y T F I V R A V D S N G A E S G N S
Minc18543b TL SG PY IQNF YKNDL NN FHT K I L F T T Y L V T G L N P N T I Y T F T V R A V D A N G A E S G N S
Minc_AAM28240 TL SG PY IQNF YKNDL NN FHT K I L F T T Y L V T G L N P N T I Y T F T V R A V D A N G A E S G N S

Mh10g200708_Contig1886:4771..7593 AEVVHQTAPNYKKXXVFDVSKLGAKGDGKTLNTKIIQNAIDKCSNNSKSPFDCKI
Mh10g200708_Contig1443:9536..11966 NQIVAKTANNYEKXXIVDITTVGAIGDGTTLNTQTIQKAIDLCSNNST-SPFGCKV
Minc18543b NQVVVKTAENYGKXXIVDITTFGATGDGTTLNTQAIQKAIDSCSSST-SAFGCKV
Minc_AAM28240 NQVVVKTAENYGKXXIVDITTFGATGDGTTLNTQAIQKAIDSCSSST-SAFGCKV

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

230 240 250 260 270

L I S A D K C G R I F L S G A L Y L R S N M T F E I A N G T I L R G S S F V D Y P L I K N V V A A L L Y G I
L I P K G - - - I F L S G P L F L R S Q M T F E L A N G A I L R A T S N A A K Y P L Q Y G - S T P S A F F N
L I P K G - - - I F L S G P L F L R S Q M T F E L A N G A I L R A T S N P S K F P N Q Y G - N T P S A F L N
L I P K G - - - I F L S G P L F L R S Q M T F E L A N G A I L R A T S N P S K F P N Q Y G - N T P S A F L N

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

280 290 300 310 320

D H X X - - I K N V R I I G F G T I D G Q Q W Q O S S S D T I D E L G N N L A Y Y E N S Y F S T W N L T G I A
A Y X X A - - - I N N I R V V G P G T I D G Q Q W Q O S S S D T I D E L G N N L A Y Y E N S Y F S T W N L T G I A
A L X X N G - S L T N I R V I G P G S V D G N G W K L A S N A I D E L G R Q I P V Y A K G S P S T V N N L G I
A L X X N G - S L T N I R V I G P G S V D G N G W K L A S N A I D E L G R Q I P V Y A K G S P S T V N N L G I

Mh10g200708_Contig1886:4771..7593 340 350 360 370 380
 Mh10g200708_Contig1443:9536..11966
 Minc18543b
 Minc_AAM28240

RSQMOAAVDRANGNISSTDLSNAYASVRSSLITFYAATNIFV-GNIRLINPAYHG LAANQIMA-----NGN-----NYVS-RSRLFAINSVSNLYIGGAITFLNPSM LAANQVQT-----HGN-----NYYS-RSRLANFNFVTNLHIGGGITFINPSM LAANQVQT-----HGN-----NYYS-RSRLANFNFVTNLHIGGGITFINPSM

Mh10g200708_Contig1886:4771..7593 390 400 410 420 430
 Mh10g200708_Contig1443:9536..11966
 Minc18543b
 Minc_AAM28240

IRFLLSRNV--XXVCAYTRTETYNINNAADGIQ-SQSINTIIFNNFLNSGDDCVVF TTLGFGDSKNVXXSIIINVRFQTYNIINNGDGIDIGRSSNIQIIGSFFDTGDDCIAI TTIGLADSKNVXXSIIISVRFQTYNIINNGDGIDIGRSSNIQIIGSFFDTGDDCIAI TTIGLADSKNVXXSIIISVRFQTYNIINNGDGIDIGRSSNIQIIGSFFDTGDDCIAI

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

450 460 470 480 490

KAGQCQAVENFPPTQYGWVFNFMREGHGVVVLGSETGSWIKELAEXXDNVAFL
 GTGCGINAGQSPPVQCILIKNNYFRHGHGAPSFGSNTGDWVKDVLIEEXXDNIAFL
 GTGCGSNAGQQGAPVQCILIKNNYFRHGHGAPAFGGSAGDGIKDVLVEEXXDNVAFL
 GTGCGSNAGQQGAPVQCILIKNNYFRHGHGAPAFGGSAGDGIKDVLVEEXXDNVAFL

Mh10g200708_Contig1886:4771..7593
Mh10g200708_Contig1443:9536..11966
Minc18543b
Minc_AAM28240

500 510 520 530 540

TENGMRVKSTAQTGGGCQNFYFRDSAMREIGTKNIEINGHLFNRNSNVLYQSGSQ
 TDNGIRLKSSPQCAGGGVQNVYVARDIAMLSVGSRNNFTFGGQQFSGDTT-----SG
 TDNGIRFKSSPQCAGGGAQNVYARDIAMQSVGSYNNFTFGGRQFSGDTT-----AG
 TDNGIRFKSSPQCAGGGAQNVYARDIAMQSVGSYNNFTFGGRQFSGDTT-----AG

Mh10g200708_Contig1886:4771..7593 CAFVLSMTYSLGSTNWVRAKTPSYFLDINVKSVTVDNGNTATGGPMISVTGYSGN
Mh10g200708_Contig1443:9536..11966 HPFVFMLNYRTTSIG--NAKIPPTQFSNITCTRISIDDNVKPTKCGGSFIYLIIGHDGG
Minc18543b HPFVFMLDYDSNPSG--NAKIPAQFKDITITRCSVDNIKPTKSGEILYAVGHDDGG
Minc_AAM28240 HPFVFMLDYDSNPSG--NAKIPAQFKDITITRCSVDNIKPTKSGEILYAVGHDDGG

Mh10g200708_Contig1886:4771..7593 ISLGYPETFSKNVXXTFENIKIIRARPANISRLQDSLFKNILIEDWGNFQSPWIL
Mh10g200708_Contig1443:9536..11966 GI--YQTXXK----FSNIKVTNAAAPAQISLADTVVFNNVDFTNYGP-NNAWSI
Minc18543b NI--YQPVYNKNIXXTFKDIKVINAAPAQIKLLDTGIFNKFDFTNFGT-NDAWSI
Minc_AAM28240 NI--YQPVYNKNIXXTFKDIKVINAAPAQIKLLDTGIFNKFDFTNFGT-NDAWSI

Mh10g200708_Contig1886:4771..7593 NNT E NL K F V N V Q P M - P N - - - -
Mh10g200708_Contig1443:9536..11966 N K A E N V K F I N V P T M K L N K L N Y A
Minc18543b T K S K G V Q F I N V P T M K L N K L N F A
Minc_AAM28240 T K S K G V Q F I N V P T M K L N K L N F A

670 680

B) PL3

G.pal_sca02986
 G.pal_sca02674
 G.rost_AAF80747
 Minc14254
 Minc11772
 Mh10g200708_Contig711:8186..6817
 Mh10g200708_Contig338:101858..103127
 Mh10g200708_Contig338:80243..81506
 Minc11928a
 Minc01517
 Minc12947
 Mh10g200708_Contig418:1229..2522
 Mh10g200708_Contig418:19714..18783
 B.muc_BAE48375
 G.pal_ctg206645
 G.rost_ACO37135
 Minc03618
 Minc04661
 Mh10g200708_Contig400:8836..9978
 Mh10g200708_Contig1669:8502..9658
 Minc18743
 Mh10g200708_Contig1669:7002..6207
 Mh10g200708_Contig331:11376..12399
 Mh10g200708_Contig841:11519..12588
 Mh10g200708_Contig841:8801..9875

The sequence alignment shows a highly conserved motif at positions 85-87 (RKN) highlighted in red. A red bracket labeled "RKN" points to this motif.

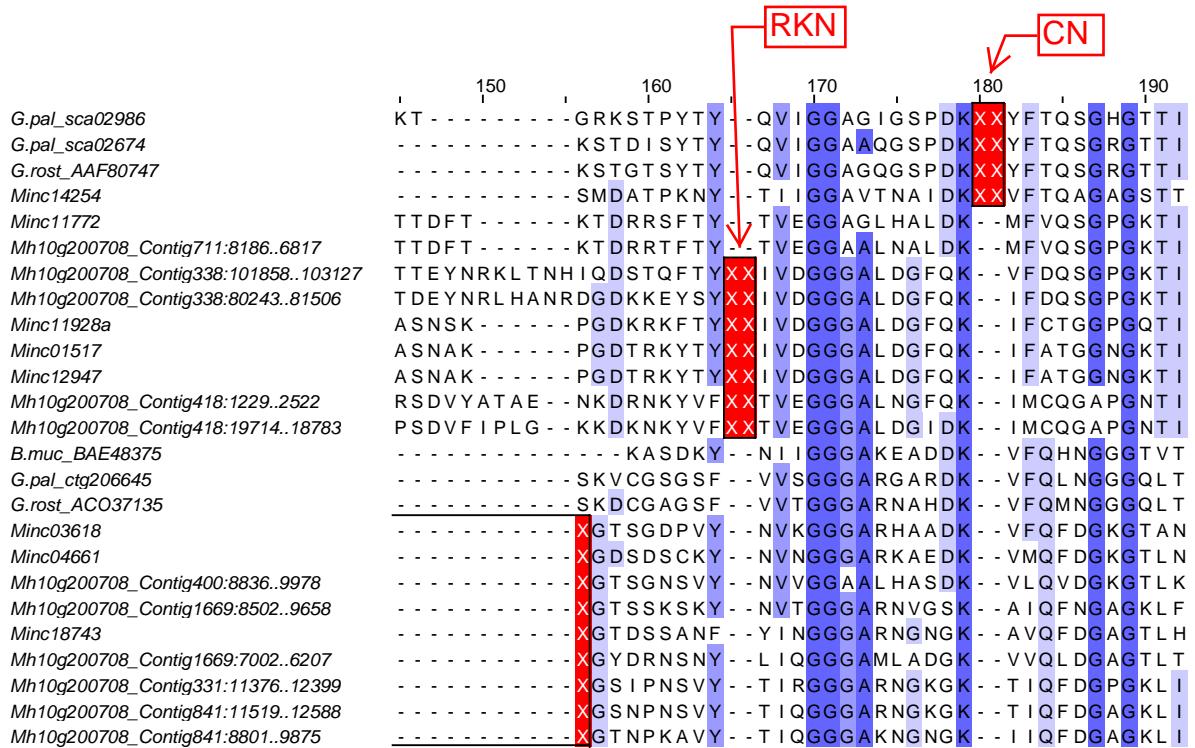
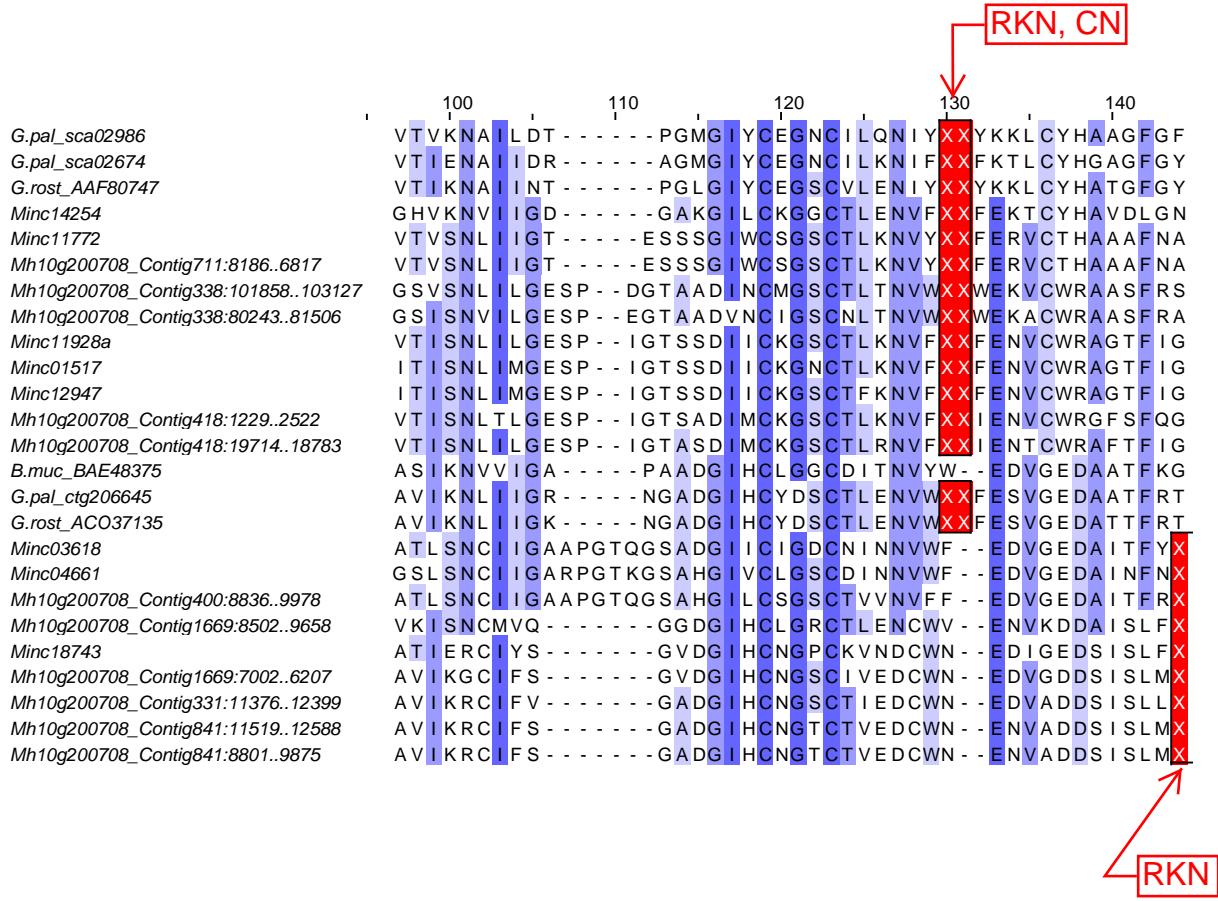
RKN

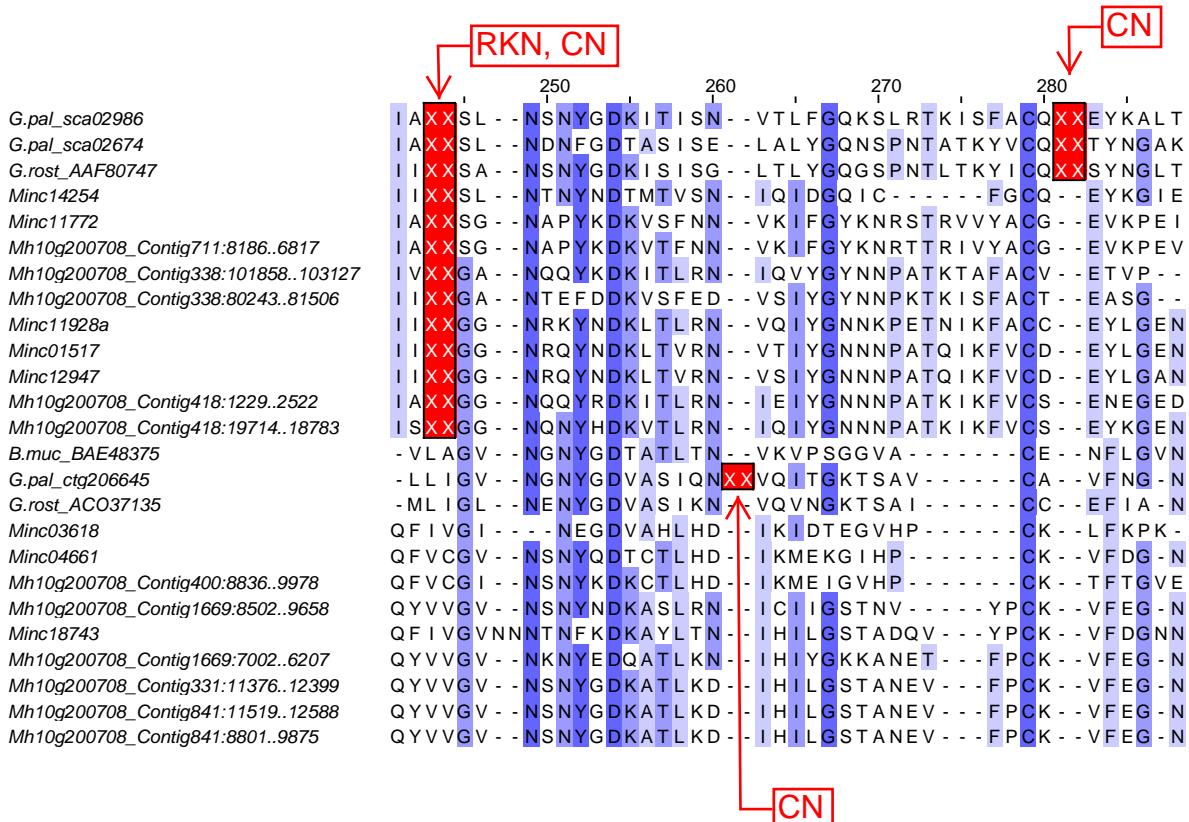
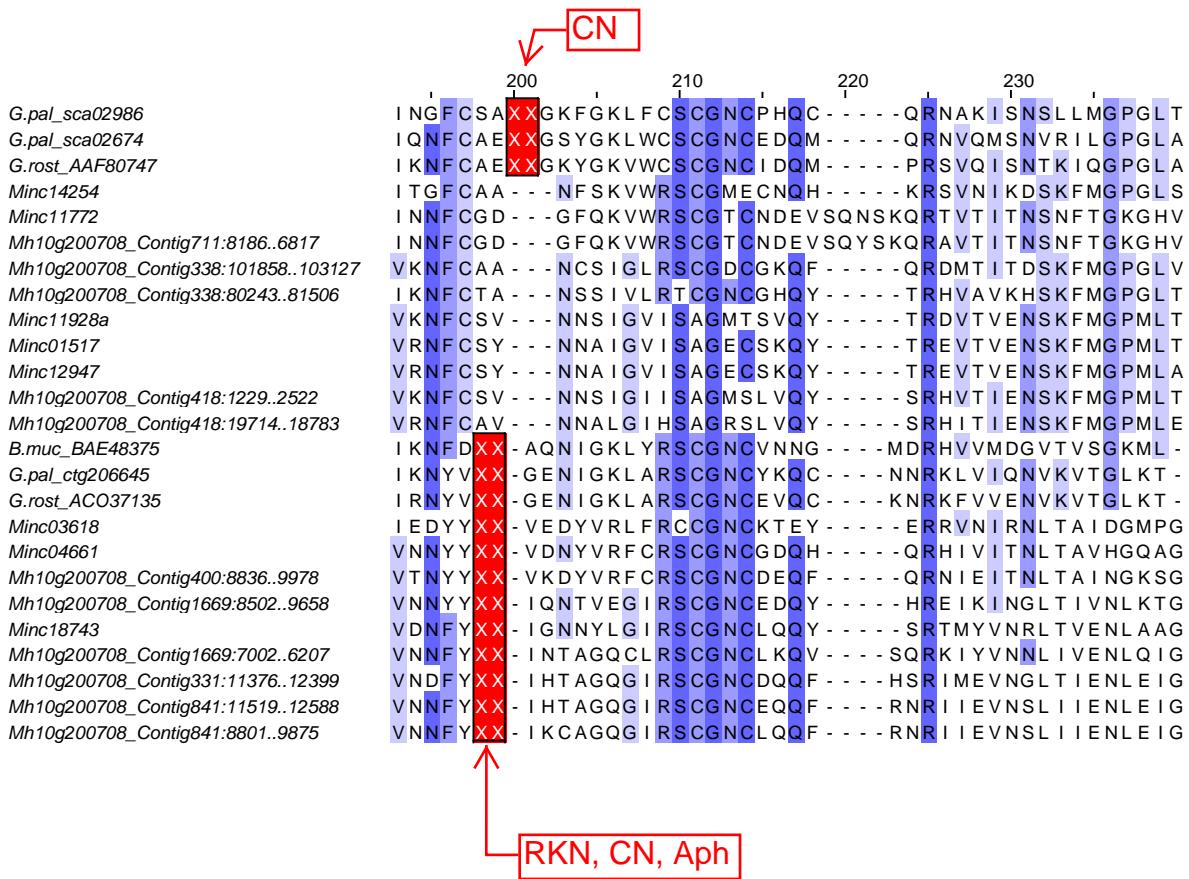
G.pal_sca02986
 G.pal_sca02674
 G.rost_AAF80747
 Minc14254
 Minc11772
 Mh10g200708_Contig711:8186..6817
 Mh10g200708_Contig338:101858..103127
 Mh10g200708_Contig338:80243..81506
 Minc11928a
 Minc01517
 Minc12947
 Mh10g200708_Contig418:1229..2522
 Mh10g200708_Contig418:19714..18783
 B.muc_BAE48375
 G.pal_ctg206645
 G.rost_ACO37135
 Minc03618
 Minc04661
 Mh10g200708_Contig400:8836..9978
 Mh10g200708_Contig1669:8502..9658
 Minc18743
 Mh10g200708_Contig1669:7002..6207
 Mh10g200708_Contig331:11376..12399
 Mh10g200708_Contig841:11519..12588
 Mh10g200708_Contig841:8801..9875

The sequence alignment shows two motifs highlighted in red: RKN (around position 55) and CN (around position 65). Red arrows point to each of these motifs.

RKN

CN

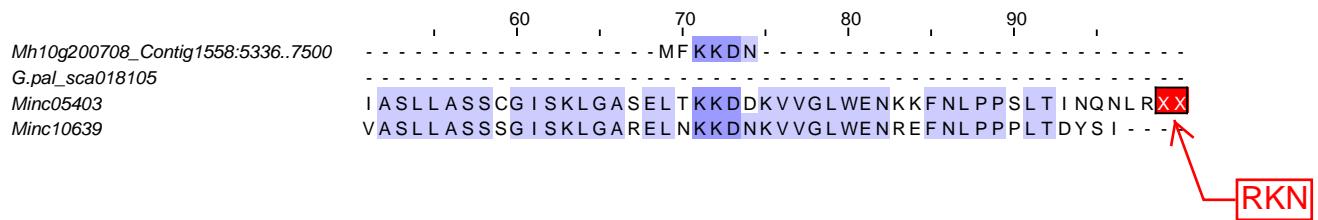
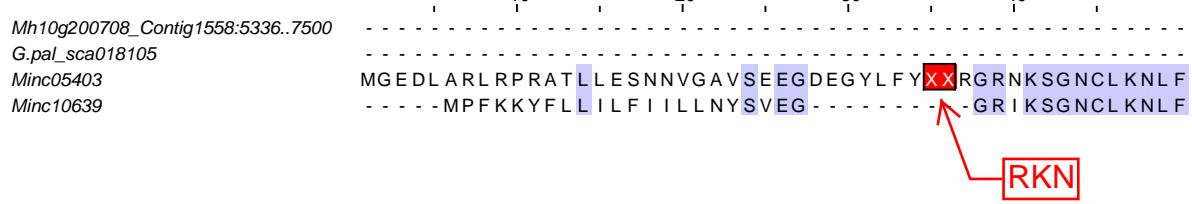




	290	300	310	320	330
<i>G.pal_sca02986</i>	- - -	- - -	- - -	- - -	- - -
<i>G.pal_sca02674</i>	- - -	- - -	- - -	- - -	- - -
<i>G.rost_AAF80747</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc14254</i>	R S G	- - -	- - -	- - -	- - -
<i>Minc11772</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig711:8186..6817</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig338:101858..103127</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig338:80243..81506</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc11928a</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc01517</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc12947</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig418:1229..2522</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig418:19714..18783</i>	- - -	- - -	- - -	- - -	- - -
<i>B.muc_BAE48375</i>	- - -	- - -	- - -	- - -	- - -
<i>G.pal_ctg206645</i>	- - -	- - -	- - -	- - -	- - -
<i>G.rost_ACO37135</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc03618</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc04661</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig400:8836..9978</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig1669:8502..9658</i>	- - -	- - -	- - -	- - -	- - -
<i>Minc18743</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig1669:7002..6207</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig331:11376..12399</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig841:11519..12588</i>	- - -	- - -	- - -	- - -	- - -
<i>Mh10g200708_Contig841:8801..9875</i>	- - -	- - -	- - -	- - -	- - -

<i>G.pal_sca02986</i>	- - -
<i>G.pal_sca02674</i>	- - -
<i>G.rost_AAF80747</i>	- - -
<i>Minc14254</i>	R S G
<i>Minc11772</i>	- - -
<i>Mh10g200708_Contig711:8186..6817</i>	- - -
<i>Mh10g200708_Contig338:101858..103127</i>	- - -
<i>Mh10g200708_Contig338:80243..81506</i>	- - -
<i>Minc11928a</i>	- - -
<i>Minc01517</i>	- - -
<i>Minc12947</i>	- - -
<i>Mh10g200708_Contig418:1229..2522</i>	- - -
<i>Mh10g200708_Contig418:19714..18783</i>	- - -
<i>B.muc_BAE48375</i>	- - -
<i>G.pal_ctg206645</i>	- - -
<i>G.rost_ACO37135</i>	- - -
<i>Minc03618</i>	- - -
<i>Minc04661</i>	A K N
<i>Mh10g200708_Contig400:8836..9978</i>	K - -
<i>Mh10g200708_Contig1669:8502..9658</i>	- - -
<i>Minc18743</i>	- - -
<i>Mh10g200708_Contig1669:7002..6207</i>	- - -
<i>Mh10g200708_Contig331:11376..12399</i>	- - -
<i>Mh10g200708_Contig841:11519..12588</i>	- - -
<i>Mh10g200708_Contig841:8801..9875</i>	- - -

C) GH43

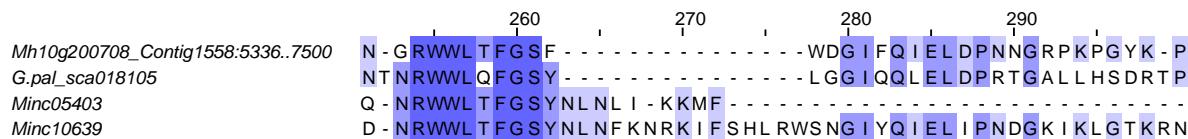
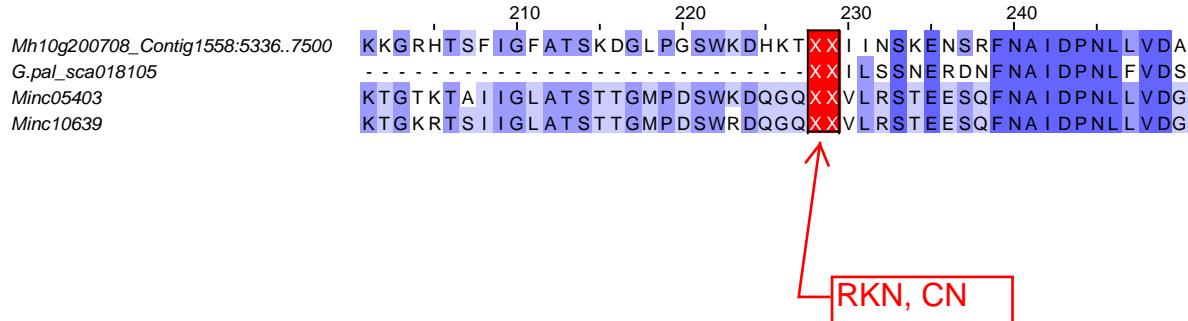


Mh10g200708_Contig1558:5336..7500
G.pal_sca018105
Minc05403
Minc10639

	110	120	130	140
<i>Mh10g200708_Contig1558:5336..7500</i>	- X K L K G D Y R V H D P S I I Q L R N K S F I I Y S T H N G I E A R I S N N L I N W K R S S -			
<i>G.pal_sca018105</i>	- - - - - M V Q L S N G L F V L Y S T H N G M E E R Q S T D L A H W Y R A S K			
<i>Minc05403</i>	A Q S H K I S G N I T L H D P C I A Q L K N G N F I I Y S T H N G L E A R I S S D L F E W E G A G -			
<i>Minc10639</i>	- - K G N X X G N I T L H D P C I A Q L K N G N F I I Y S T H N G L E A R I S S D L F E W E G A G -			

Mh10g200708_Contig1558:5336..7500
G.pal_sca018105
Minc05403
Minc10639

	160	170	180	190
<i>Mh10g200708_Contig1558:5336..7500</i>	- - S A F P N G L L Q W A N - S L T K D W N E F W A P D I S F H N N K - - - - F W L Y Y S - - V P Y			
<i>G.pal_sca018105</i>	- - S A F P N G L L A W A N X X - S L T K D P N E F W A P D V S G P H G I G K K K K Y W L F Y A - - V P -			
<i>Minc05403</i>	- - F A F S Q G V P W A R - G L T K D W N E L W A P D I S M H G G L - - - - Y W L Y Y A X X V P V			
<i>Minc10639</i>	- - F A F S Q G V P W A R - G L T K D W N E L W A P D I S I H G G V - - - - Y W L Y Y A X X V P V			



Mh10g200708_Contig1558:5336..7500 310 320 330 340
G.pal_sca018105
Minc05403
Minc10639

```

F N I A K R S R K F M R A I E A P F V F Y K X X K G F Y Y L F V S F D K C C S G L Q S T Y S I H V G
Q V I A A R P D V E D H A I E A P F L F R K G X X K W H F L F V S F G H C C Q G T A S T Y S I H V G
H L A A R - - - D A G I E A P F I I H R X X G N F Y Y L F V S F G T C C A G L Q S T Y S I H V G

```

RKN, CN

Mh10g200708_Contig1558:5336..7500 360 370 380 390
G.pal_sca018105
Minc05403
Minc10639

```

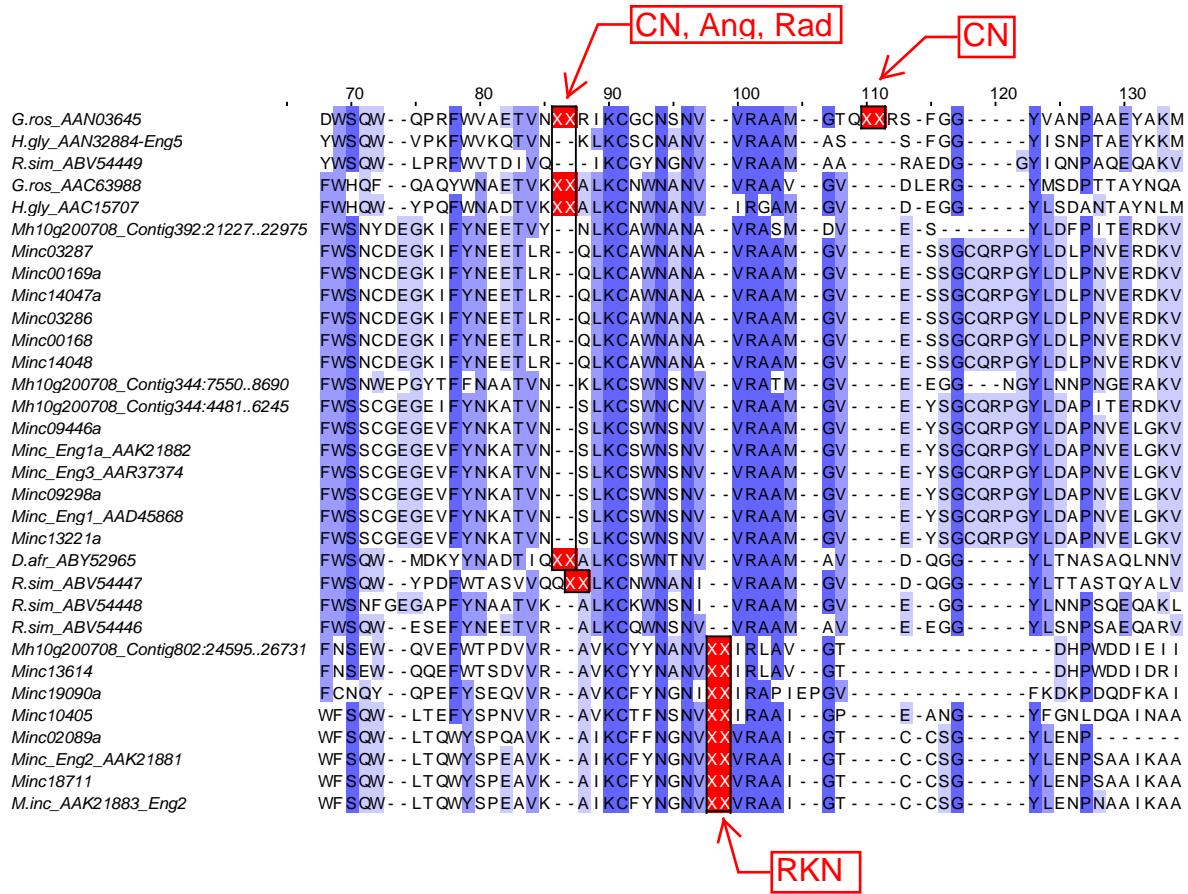
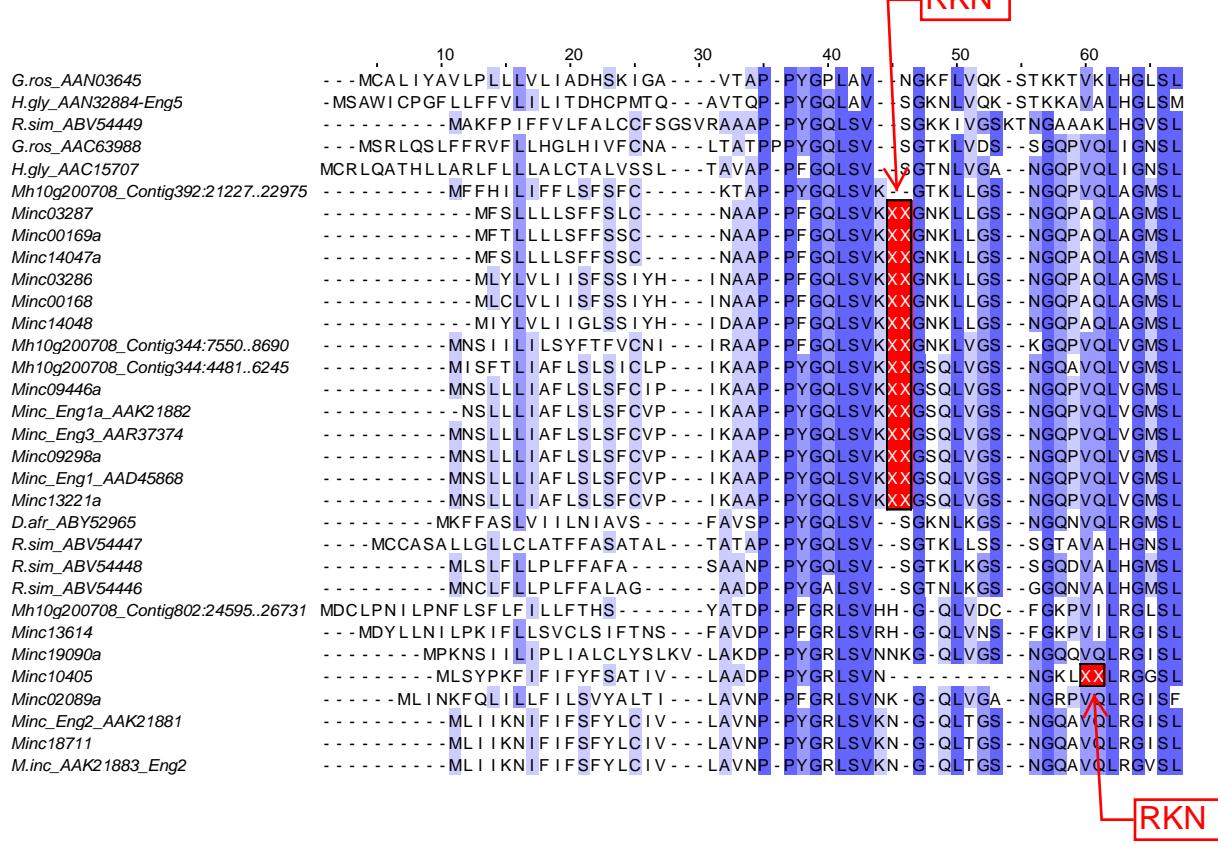
R S E S P N G P F V D D K N I K M M D G G G M L L L G S N G I E I G P G Q Q S V Y K A K G F G D -
R S D S A N G P F R D R R G I L M T A G G G T L L L G T N G P E V G P G Q Q S V V T T E D K R D -
R S H R P F G P Y L D D K N V P M L Q G G G M P L L S S N N Q K I G P G Q Q S L L K I K R K G K K N

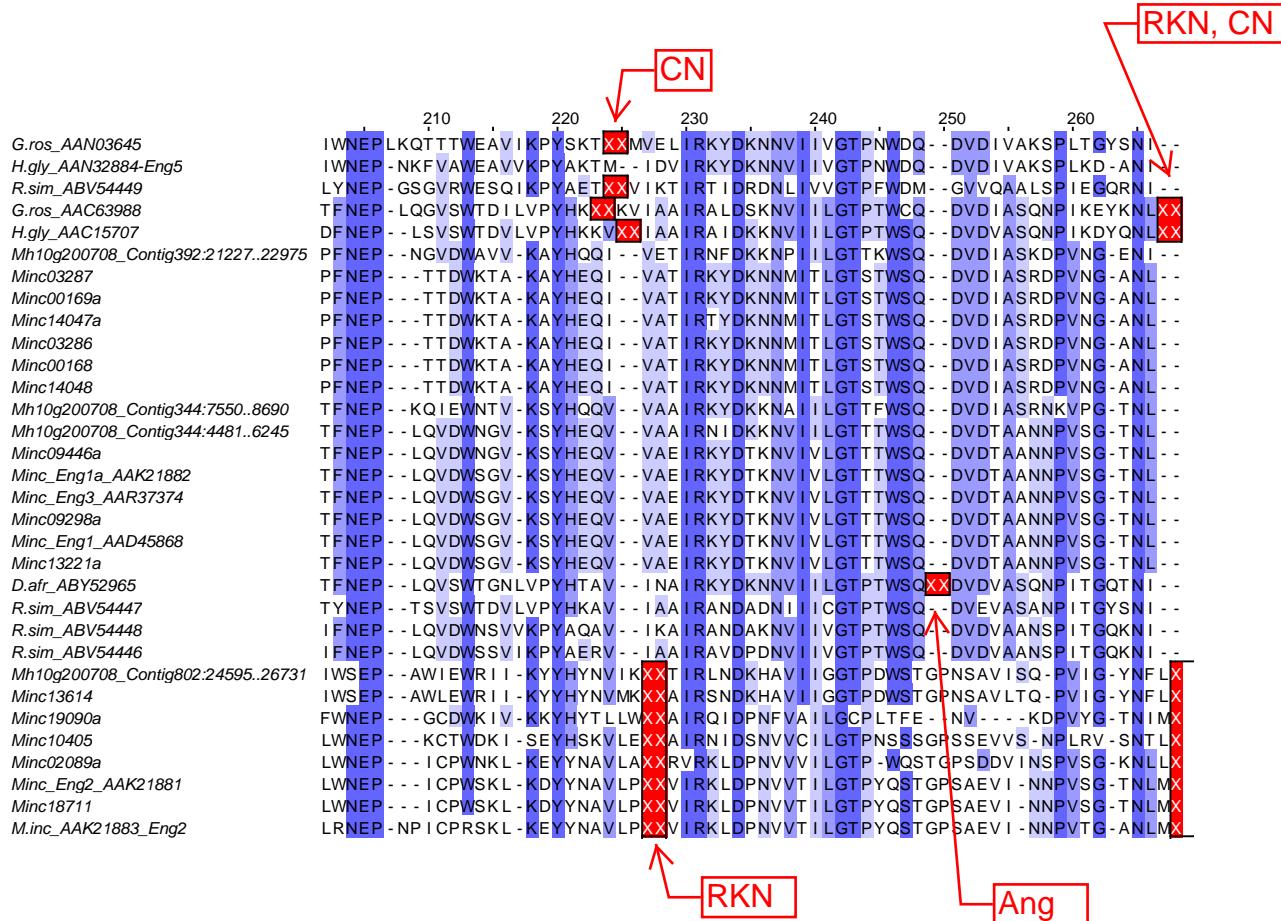
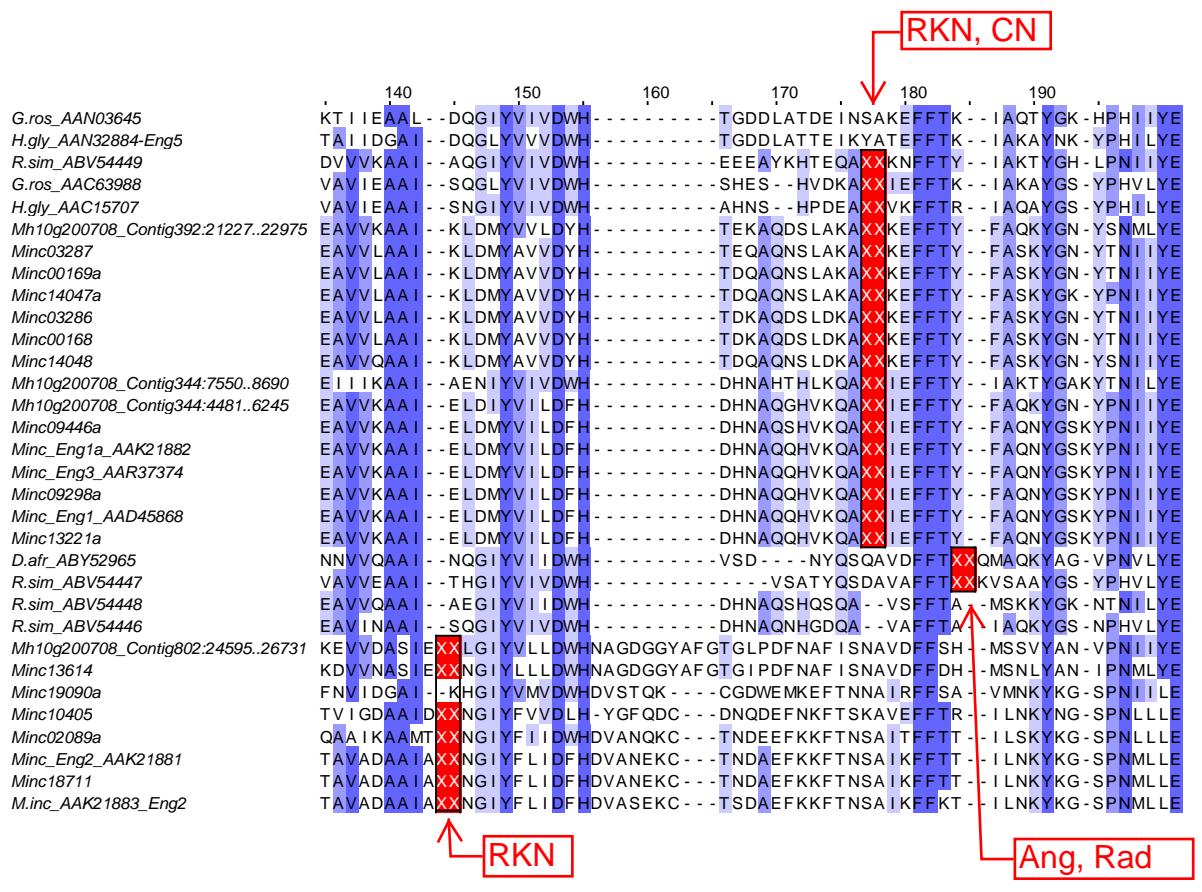
```

Mh10g200708_Contig1558:5336..7500 - - L | Y H Y Y D K T D N G M P K L G I K K I R W E N N - W P V L S E F -
G.pal_sca018105 - - I L L Y H Y Y D A N A D G R N T L G L R Q I R W D A E R W P T V - - - - -
Minc05403 - - - - -
Minc10639 M I I L V Y H Y Y D G L D N G L P K L G I K R L G W T S D G W P F V K D L Q

410 420 430

D) GH5_cel





RKN, CN, Rad

Sequence alignment showing a conserved region from position 270 to 330. The alignment is color-coded by conservation: blue for low conservation, purple for moderate conservation, and red for high conservation. A red arrow points to position 300, highlighting a specific sequence motif.

Conservation scale: 0 (blue), 1 (purple), 2 (red).

Proteins listed:

- G.ros_AAN03645
- H.gly_AAN32884-Eng5
- R.sim_ABV54449
- G.ros_AAC63988
- H.gly_AAC15707
- Mh10g200708_Contig392.21227..22975
- Minc03287
- Minc0169a
- Minc14047a
- Minc03286
- Minc00168
- Minc14048
- Mh10g200708_Contig344:7550..8690
- Mh10g200708_Contig344:4481..6245
- Minc09446a
- Minc_Eng1_AAK21882
- Minc_Eng3_AAR37374
- Minc09298a
- Minc_Eng1_AAD45868
- Minc13221a
- D.afr_ABY52965
- R.sim_ABV54447
- R.sim_ABV54448
- R.sim_ABV54446
- Mh10g200708_Contig802:24595..26731
- Minc13614
- Minc19090a
- Minc10405
- Minc02089a
- Minc_Eng2_AAK21881
- Minc18711
- M.inc_AAK21883_Eng2

RKN

CN, RKN, Rad

Sequence alignment showing a conserved region from position 340 to 400. The alignment is color-coded by conservation: blue for low conservation, purple for moderate conservation, and red for high conservation. A red arrow points to position 370, highlighting a specific sequence motif.

Conservation scale: 0 (blue), 1 (purple), 2 (red).

Proteins listed:

- G.ros_AAN03645
- H.gly_AAN32884-Eng5
- R.sim_ABV54449
- G.ros_AAC63988
- H.gly_AAC15707
- Mh10g200708_Contig392.21227..22975
- Minc03287
- Minc0169a
- Minc14047a
- Minc03286
- Minc00168
- Minc14048
- Mh10g200708_Contig344:7550..8690
- Mh10g200708_Contig344:4481..6245
- Minc09446a
- Minc_Eng1_AAK21882
- Minc_Eng3_AAR37374
- Minc09298a
- Minc_Eng1_AAD45868
- Minc13221a
- D.afr_ABY52965
- R.sim_ABV54447
- R.sim_ABV54448
- R.sim_ABV54446
- Mh10g200708_Contig802:24595..26731
- Minc13614
- Minc19090a
- Minc10405
- Minc02089a
- Minc_Eng2_AAK21881
- Minc18711
- M.inc_AAK21883_Eng2

Ang

	410	420	430	440	450	460	
<i>G.ros_AAN03645</i>	-	-	-	-	-	-	-
<i>H.gly_AAN32884-Eng5</i>	RKNE	-	-	-	-	-	-
<i>R.sim_ABV54449</i>	STSSGSSSSNSGNSAATTTKKPPS	-	-	-	-	-	-
<i>G.ros_AAC63988</i>	SSSGSSSSGSSG-	-	-	-	-	-	-
<i>H.gly_AAC15707</i>	NNGKYP\$GD - ITTKLPSKSNNA	-	-	-	-	-	-
<i>Mh10g200708_Contig392.21227..22975</i>	GKQNNAAPSPS-	-	-	-	-	-	-
<i>Minc03287</i>	GNPNNAAPSPSNNN	-	-	-	-	-	-
<i>Minc00169a</i>	NKPNNAAAPSPSNNN	-	-	-	-	-	-
<i>Minc14047a</i>	SKGKYP\$GDE I ITTKSPSKNKYPGDEVPSKTS-SNTKYPGDNIIITTKLPSKTN	-	-	-	-	-	-
<i>Minc03286</i>	NKGKYP\$GDD II ITTKSPSKNKYPGDEVPSKTRPNNNYPGNNII LTTKSPSNGNNNNKYP\$AP	-	-	-	-	-	-
<i>Minc00168</i>	NKGKYP\$GDD II ITTKSPSKNKYPGDEVPSKTRPNNNYPGNNII LTTKSPSNGNNNNKYP\$AP	-	-	-	-	-	-
<i>Minc14048</i>	NKGKYP\$GDD II ITTKSPSKNKYPGDEVPSKTRPNNNYPGNNII LTTKSPSNGNNNNKYP\$AP	-	-	-	-	-	-
<i>Mh10g200708_Contig344:7550..8690</i>	SSSSGSSSSNNNANNNSGNSNNNGNSKTVNI	-	-	-	-	-	-
<i>Mh10g200708_Contig344:4481..6245</i>	ASSSSGSGSKPSGNKPSNSESSTA TKTSGNSGN	-	-	-	-	-	-
<i>Minc09446a</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>Minc_Eng1_a_AAK21882</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>Minc_Eng3_AAR37374</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>Minc09298a</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>Minc_Eng1_AAD45868</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>Minc13221a</i>	ASPGSGGSNPSPGNKPSNSQTAKTSSNSGN	-	-	-	-	-	-
<i>D.afr_ABY52965</i>	GSGSTATTKPPTGQTTT	-	-	-	-	-	-
<i>R.sim_ABV54447</i>	SATTTTTTKSGATTTTT	-	-	-	-	-	-
<i>R.sim_ABV54448</i>	SAATTTTTTRAT	-	-	-	-	-	-
<i>Mh10g200708_Contig802:24595..26731</i>	-	-	-	-	-	-	-
<i>Minc13614</i>	-	-	-	-	-	-	-
<i>Minc19090a</i>	-	-	-	-	-	-	-
<i>Minc10405</i>	-	-	-	-	-	-	-
<i>Minc02089a</i>	-	-	-	-	-	-	-
<i>Minc_Eng2_AAK21881</i>	-	-	-	-	-	-	-
<i>Minc18711</i>	-	-	-	-	-	-	-
<i>M.inc_AAK21883_Eng2</i>	-	-	-	-	-	-	-

	480	490	500	510	520	530	
<i>G.ros_AAN03645</i>	-	-	-	-	-	-	-
<i>H.gly_AAN32884-Eng5</i>	-	-	-	-	-	-	-
<i>R.sim_ABV54449</i>	-	-	-	-	-	-	-
<i>G.ros_AAC63988</i>	-	-	-	-	-	-	-
<i>H.gly_AAC15707</i>	-	-	-	-	-	-	-
<i>Mh10g200708_Contig392.21227..22975</i>	-	-	-	-	-	-	-
<i>Minc03287</i>	-	-	-	-	-	-	-
<i>Minc00169a</i>	-	-	-	-	-	-	-
<i>Minc14047a</i>	-	-	-	-	-	-	-
<i>Minc03286</i>	-	-	-	-	-	-	-
<i>Minc00168</i>	-	-	-	-	-	-	-
<i>Minc14048</i>	-	-	-	-	-	-	-
<i>Mh10g200708_Contig344:7550..8690</i>	-	-	-	-	-	-	-
<i>Mh10g200708_Contig344:4481..6245</i>	-	-	-	-	-	-	-
<i>Minc09446a</i>	-	-	-	-	-	-	-
<i>Minc_Eng1_a_AAK21882</i>	-	-	-	-	-	-	-
<i>Minc_Eng3_AAR37374</i>	-	-	-	-	-	-	-
<i>Minc09298a</i>	-	-	-	-	-	-	-
<i>Minc_Eng1_AAD45868</i>	-	-	-	-	-	-	-
<i>Minc13221a</i>	-	-	-	-	-	-	-
<i>D.afr_ABY52965</i>	-	-	-	-	-	-	-
<i>R.sim_ABV54447</i>	-	-	-	-	-	-	-
<i>R.sim_ABV54448</i>	-	-	-	-	-	-	-
<i>R.sim_ABV54446</i>	-	-	-	-	-	-	-
<i>Mh10g200708_Contig802:24595..26731</i>	-	-	-	-	-	-	-
<i>Minc13614</i>	-	-	-	-	-	-	-
<i>Minc19090a</i>	-	-	-	-	-	-	-
<i>Minc10405</i>	-	-	-	-	-	-	-
<i>Minc02089a</i>	-	-	-	-	-	-	-
<i>Minc_Eng2_AAK21881</i>	-	-	-	-	-	-	-
<i>Minc18711</i>	-	-	-	-	-	-	-
<i>M.inc_AAK21883_Eng2</i>	-	-	-	-	-	-	-

G.ros_AAN03645
 H.gly_AAN32884-Eng5
 R.sim_ABV54449
 G.ros_AAC63988
 H.gly_AAC15707
 Mh10g200708_Contig392.21227..22975
 Minc03287
 Minc00169a
 Minc14047a
 Minc03286
 Minc00168
 Minc14048
 Mh10g200708_Contig344:7550..8690
 Mh10g200708_Contig344:4481..6245
 Minc09446a
 Minc_Eng1a_AAK21882
 Minc_Eng3_AAR37374
 Minc09298a
 Minc_Eng1_AAD45868
 Minc13221a
 D.afr_ABY52965
 R.sim_ABV54447
 R.sim_ABV54448
 R.sim_ABV54446
 Mh10g200708_Contig802:24595..26731
 Minc13614
 Minc19090a
 Minc10405
 Minc02089a
 Minc_Eng2_AAK21881
 Minc18711
 M.inc_AAK21883_Eng2

```

      540   550   560   570   580   590   600
      - - - - GSSGSSSGSSSG - - - - TSTKKPNNNNNGKYRSGDVT ITKS - - - -
      - - - - TKTDTTKNNS - - - - TKTDPNKNNS - - - -
      - - - - TKTDPNKNNS - - - -
      - - - - DAVPATKSPSKNNNNNGKYPGDNV I LTT I SPSNNNKYPSPG - - - - PSNNNNKYPGSD
      GPSKDNNNGKSPSAPS KDNNNNGKYPGDN I I ITTKSPS - KNKYPGDDVPATKSPSKNNNNKYPGSD
      GPSKDNNNGKSPSAPS KDNNNNGKYPGDN I I ITTKSPS - -
      - - - - SSGSNTGSNSGNTGSN - - - -
      - - - - SNTXXGNTGSNAGANSGNTGTS - - - -
      - - - - SNTXXGNTGSNAGANSGNTGTS - - - -
      - - - - SNTXXGNTGSNAGANSGNTGTS - - - -
      - - - - SNTXXGNTGSNAGASSGNTGTS - - - -
      - - - - SNTXXGNTGSNAGASSGNTGTS - - - -
      - - - - PTTTTRAAASSG - - - -
      - - - - RKN
  
```

G.ros_AAN03645
 H.gly_AAN32884-Eng5
 R.sim_ABV54449
 G.ros_AAC63988
 H.gly_AAC15707
 Mh10g200708_Contig392.21227..22975
 Minc03287
 Minc00169a
 Minc14047a
 Minc03286
 Minc00168
 Minc14048
 Mh10g200708_Contig344:7550..8690
 Mh10g200708_Contig344:4481..6245
 Minc09446a
 Minc_Eng1a_AAK21882
 Minc_Eng3_AAR37374
 Minc09298a
 Minc_Eng1_AAD45868
 Minc13221a
 D.afr_ABY52965
 R.sim_ABV54447
 R.sim_ABV54448
 R.sim_ABV54446
 Mh10g200708_Contig802:24595..26731
 Minc13614
 Minc19090a
 Minc10405
 Minc02089a
 Minc_Eng2_AAK21881
 Minc18711
 M.inc_AAK21883_Eng2

```

      610   620   630   640   650   660
      - - - - PSNNNNNGKYP - - - -
      - - - - DVTTTKSP - - SNNKYP SAPS NGNNNGKYPGEN V LTTKSPSND NKYP SAP SGPS KDNNGECPGD
      N I I ITTKSPSNNNKYP SAPSN - NNNKYPGDN V I LTTKSPSND NKYP SAP SGPS NDNN GECPSED
      - - - - NNNKYP SAPSN - NNNKYPGDN V I LTTKSPSND NKYP SAP SGPS NDNN GECPSED
  
```

	680	690	700	710	720	730	
G.ros_AAN03645	-	-	-	-	-	-	-
H.gly_AAN32884-Eng5	-	-	-	-	-	-	-RYF
R.sim_ABV54449	-	-	-	-	-	-	-SGSASASVTVVSTNTWN
G.ros_AAC63988	-	-	-	-	-	-	-SGSASISVVPSEN-TWN
H.gly_AAC15707	-	-	-	-	-	-	-SGDVSLSIELRD-KWV
Mh10g200708_Contig392.21227..22975	-	-	-	-	-	-	-KGGGKISISLNAEQTWP
Minc03287	-	-	-	-	-	-	-KGGGKISISLTADQTWP
Minc00169a	-	-	-	-	-	-	-KGGGKISISLTADQTWP
Minc14047a	-	-	-	-	-	-	-DIIITTK--SHNNKYPGPS-NNNKYPGSDQVTTQSPSNDNNNGNSSHVNNEQQCSISLTVEENGSWV
Minc03286	-	-	-	-	-	-	-DIIITTKLPDNNKYPGPSNNNNKYPAGGQVTTIQSPSNDNNNGNSSHGEQQCSISLTVEENGSWV
Minc00168	-	-	-	-	-	-	-DIIITTKLPDNNKYPGPSNNNNKYPAGGQVTTIQSPSNDNNNGNSSHGEQQCSISLTVEENGSWV
Minc14048	-	-	-	-	-	-	-DIIITTKLPDNNKYPGPSNNNNKYPAGGQVTTIQSPSNDNNNGNSSHGEQQCSISLTVEENGSWV
Mh10g200708_Contig344.7550..8690	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Mh10g200708_Contig344.4481..6245	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Minc09446a	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Minc_Eng1_a_AAK21882	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Minc_Eng3_AAR37374	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Minc09298a	-	-	-	-	-	-	-TGSSSVTASVQVPDKWD
Minc_Eng1_AAD45868	-	-	-	-	-	-	-TSGSSVTASVQVPDKWD
Minc13221a	-	-	-	-	-	-	-TSGSSVTASVQVPDKWD
D.afr_ABV52965	-	-	-	-	-	-	-GGNGKLVAQAQVMASSWN
R.sim_ABV54447	-	-	-	-	-	-	-SSSSSVTASVSTTNWS
R.sim_ABV54448	-	-	-	-	-	-	-SGSVSAAVSRVN-SWE
Mh10g200708_Contig802.24595..26731	-	-	-	-	-	-	-
Minc13614	-	-	-	-	-	-	-
Minc19090a	-	-	-	-	-	-	-
Minc10405	-	-	-	-	-	-	-
Minc02089a	-	-	-	-	-	-	-
Minc_Eng2_AAK21881	-	-	-	-	-	-	-
Minc18711	-	-	-	-	-	-	-
M.inc_AAK21883_Eng2	-	-	-	-	-	-	-

	740	750	760	770	780	790	800	
G.ros_AAN03645	-	-	-	-	-	-	-	-
H.gly_AAN32884-Eng5	-	-	-	-	-	-	-	-
R.sim_ABV54449	-	-	-	-	-	-	-	-
G.ros_AAC63988	-	-	-	-	-	-	-	-
H.gly_AAC15707	-	-	-	-	-	-	-	-
Mh10g200708_Contig392.21227..22975	-	-	-	-	-	-	-	-
Minc03287	-	-	-	-	-	-	-	-
Minc00169a	-	-	-	-	-	-	-	-
Minc14047a	-	-	-	-	-	-	-	-
Minc03286	-	-	-	-	-	-	-	-
Minc00168	-	-	-	-	-	-	-	-
Minc14048	-	-	-	-	-	-	-	-
Mh10g200708_Contig344.7550..8690	-	-	-	-	-	-	-	-
Mh10g200708_Contig344.4481..6245	-	-	-	-	-	-	-	-
Minc09446a	-	-	-	-	-	-	-	-
Minc_Eng1_a_AAK21882	-	-	-	-	-	-	-	-
Minc_Eng3_AAR37374	-	-	-	-	-	-	-	-
Minc09298a	-	-	-	-	-	-	-	-
Minc_Eng1_AAD45868	-	-	-	-	-	-	-	-
Minc13221a	-	-	-	-	-	-	-	-
D.afr_ABV52965	-	-	-	-	-	-	-	-
R.sim_ABV54447	-	-	-	-	-	-	-	-
R.sim_ABV54448	-	-	-	-	-	-	-	-
R.sim_ABV54446	-	-	-	-	-	-	-	-
Mh10g200708_Contig802.24595..26731	-	-	-	-	-	-	-	-
Minc13614	-	-	-	-	-	-	-	-
Minc19090a	-	-	-	-	-	-	-	-
Minc10405	-	-	-	-	-	-	-	-
Minc02089a	-	-	-	-	-	-	-	-
Minc_Eng2_AAK21881	-	-	-	-	-	-	-	-
Minc18711	-	-	-	-	-	-	-	-
M.inc_AAK21883_Eng2	-	-	-	-	-	-	-	-

CN, RKN, Rad, Ang

	810	820	
<i>G.ros_AAN03645</i>	-	-	-
<i>H.gly_AAN32884-Eng5</i>	-	-	-
<i>R.sim_ABV54449</i>	-	-	-
<i>G.ros_AAC63988</i>	GMT LNGSGK -	PTAKVLSTTA <u>C</u> SG	
<i>H.gly_AAC15707</i>	GLTFNGDKD -	PTAKIVTTKKC -	
<i>Mh10g200708_Contig392.21227..22975</i>	GLVVNGPAT -	LPQIEFLSQGVCKY	
<i>Minc03287</i>	GMGVSGSSS -	TPKISLVGSGN -	
<i>Minc00169a</i>	GMGVSGSSS -	APKISLVSSGN -	
<i>Minc14047a</i>	GMGVSGSSS -	APKISLVSSGN -	
<i>Minc03286</i>	GLVVNGPAT -	LPQIKIIGQGECKY	
<i>Minc00168</i>	GLVVNGPAT -	LPQIKVVGQGECKY	
<i>Minc14048</i>	GLVVNGPAT -	LPQIKVVGQGECKY	
<i>Mh10g200708_Contig344.7550..8690</i>	-	-	-
<i>Mh10g200708_Contig344.4481..6245</i>	GMSITGPAT -	APQIKVLGDGRCVF	
<i>Minc09446a</i>	GMNINGPAT -	PPQIKVLGDGKCVH	
<i>Minc_Eng1_a_AAK21882</i>	GMNINGPAT -	PPQIKVLGDGKCVF	
<i>Minc_Eng3_AAR37374</i>	GMNINGPAT -	PPQIKVLGDGKCVF	
<i>Minc09298a</i>	GMNINGPAT -	PPQIKVLGDGKCVF	
<i>Minc_Eng1_AAD45868</i>	GMNINGPAT -	PPQIKVLGDGKCVF	
<i>Minc13221a</i>	GMNINGPAT -	PPQIKVLGDGKCVF	
<i>D.afr_ABY52965</i>	GMSINGSNTALPTVSVVSYGYC -	-	-
<i>R.sim_ABV54447</i>	GMVLTGSGT -	PTVSIVSTSGC -	-
<i>R.sim_ABV54448</i>	GMT LNGSGT -	PTVAVAVASSSA <u>C</u> -	-
<i>Mh10g200708_Contig802.24595..26731</i>	-	-	SVY
<i>Minc13614</i>	-	-	I Y
<i>Minc19090a</i>	-	-	GC
<i>Minc10405</i>	-	-	-
<i>Minc02089a</i>	-	-	GC
<i>Minc_Eng2_AAK21881</i>	-	-	GC
<i>Minc18711</i>	-	-	-
<i>M.inc_AAK21883_Eng2</i>	-	-	-

E) GH5_xyl

	10	20	30	40	50
<i>R.sim_ABZ78968</i>	- - MF ALL LVPVFL ALVACAANVNAATGSVALGSMRQT	I QGF GGSSAWMGAMT	D AQMN		
<i>Mh_Contig2188.1417..3259</i>	MNKL K I FLL I F I I I FSSQFD C INGD I N I NLNDKRQV I DGFGACSAWQGV V SDQ I MK				
<i>Minc18739</i>	MNNL - I FLL I I FLL I I FSSQFD RNNNGNVT I NLNEKRQV I DGFGASTAWQGA V SDQ TMN				
<i>Minc11164</i>	MNNL - I FLL I I FLL I I FSSQFD RNNNGNVT I NLNEKRQV I DGFGASTAWQGA V SDQ IMN				
<i>Minc18650</i>	MNNL - I FLL I I FLL I I FSSQFD RNNNGNVT I NLNEKRQV I DGFGASTAWQGA V SDQ IMN				
<i>Minc18912</i>	MNNL - I FLL I I FLL I I FSSQFD RNNNGNVT I NLNEKRQV I DGFGASTAWQGA V SDQ IMN				
<i>Minc09178</i>	MKLFNF F LFNL I LFYYSVKCDN - I AKI NSDITYQL I DGFGSSA WLGNIPDKGIG				

	60	70	80	90	100	110
<i>R.sim_ABZ78968</i>	T L XXX FG NGNNNNQ YGL SLL RL R I DP - - - GKS WAN - - - ELS NAQ KAGARGAQVFA					
<i>Mh_Contig2188.1417..3259</i>	EL XXX YET - - - LGYS I L R I R IDE - - - NKKWNE - - - ELL NAK KALQFNAKVFA					
<i>Minc18739</i>	EL XXX FGT - - - LGYS I L R I R IDE - - - NKRWAD - - - ELS NAK KALKLNATVFA					
<i>Minc11164</i>	EL XXX FGT - - - LGYS I L R I R IDE - - - NKQWAD - - - ELS NAK KALKFNNAKVFA					
<i>Minc18650</i>	EL XXX FGT - - - LGYS I L R I R IDE - - - NKQWAD - - - ELS NAK KALKLNNAKVFA					
<i>Minc18912</i>	EL XXX FGT - - - LGYS I L RVR IDE - - - DKQWDD - - - ELS NAK KALKLNNAKVFA					
<i>Minc09178</i>	N I - - FGK - - - LGL S I L RVG I VDLCKNQKGNYRCIGQEALT A QK ASKYGVKIFA					


RKN, Rad

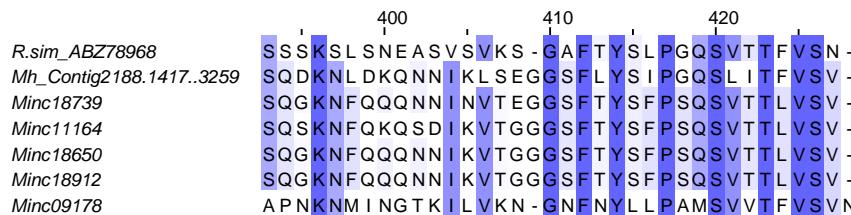
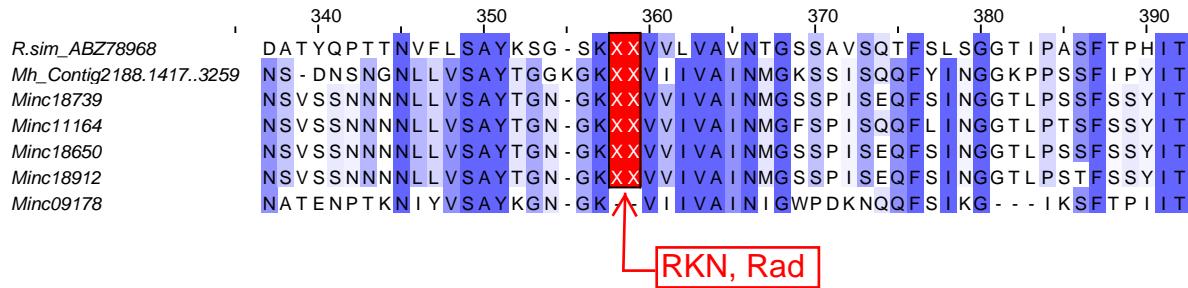
<i>R.sim_ABZ78968</i> <i>Mh_Contig2188.1417..3259</i> <i>Minc18739</i> <i>Minc11164</i> <i>Minc18650</i> <i>Minc18912</i> <i>Minc09178</i>	TPWSPPPASMKSNNNVVGGSLNTASYGAYAA ¹²⁰ SPWSAPSNNMKNNK----ALLSSQYSNYAYYL ¹³⁰ SPWTAAPANMKVNKQDEPGPLSSNQYSDYADYL ¹⁴⁰ SPWSA ¹⁵⁰ PAIMKVDNKQDEPGPLSSNQYSNYADYL SPWSA ¹⁶⁰ PAIMKDNKQDKHGPLSSNQYSNYADYL TPWTA ¹⁷⁰ PADMKDNKQDKHGPLSSNQYSNYADYL SPSTSPISFKTNYNEVMGELREDKYNDYVEYLQS ¹⁸⁰ AVDELNKVGVNLYAISLQSEPD
--	---

<i>R.sim_ABZ78968</i> <i>Mh_Contig2188.1417..3259</i> <i>Minc18739</i> <i>Minc11164</i> <i>Minc18650</i> <i>Minc18912</i> <i>Minc09178</i>	IT--VTYESC ¹⁷⁰ DWTAAQLVN ¹⁸⁰ FVKNYGS ¹⁹⁰ AV----G ²⁰⁰ T ²¹⁰ K ²²⁰ L ²³⁰ I ²⁴⁰ A ²⁵⁰ E ²⁶⁰ S ²⁷⁰ F ²⁸⁰ K ²⁹⁰ N ³⁰⁰ K ³¹⁰ ALT ³²⁰ T ³³⁰ D ³⁴⁰ P ³⁵⁰ I ³⁶⁰ L ³⁷⁰ XXS YANEAPYNTMSF ¹⁷⁰ STDEM ¹⁸⁰ KDFLKNNNAK ¹⁹⁰ I ²⁰⁰ KEGNNIK ²¹⁰ IMAPE ²²⁰ TYGYNVDMNNK ²³⁰ IL ²⁴⁰ XXS YS----DNPMT ¹⁷⁰ FTP ¹⁸⁰ DQMKNFL ¹⁹⁰ KNFA ²⁰⁰ RIKS ²¹⁰ GS ²²⁰ NIK ²³⁰ IMAPE ²⁴⁰ SYGYRRDMND ²⁵⁰ AI ²⁶⁰ IL ²⁷⁰ XXN FALEAPYNTMSFTPDQMKNFL ¹⁷⁰ KNSAK ¹⁸⁰ R ¹⁹⁰ IKSGS ²⁰⁰ STK ²¹⁰ IMATES ²²⁰ SYGYKT ²³⁰ DMND ²⁴⁰ AI ²⁵⁰ IL ²⁶⁰ XXN FALEAPYNTMSFTTDQMKNFL ¹⁷⁰ KNFAG ¹⁸⁰ K ¹⁹⁰ IKSGS ²⁰⁰ NIK ²¹⁰ IMATES ²²⁰ SYGYKT ²³⁰ DMND ²⁴⁰ AI ²⁵⁰ IL ²⁶⁰ XXN FALEAPYNTMSFTTDQMKNFL ¹⁷⁰ KNFAG ¹⁸⁰ R ¹⁹⁰ IKSGS ²⁰⁰ NIK ²¹⁰ IMAPE ²²⁰ SYGSAT ²³⁰ DMND ²⁴⁰ AI ²⁵⁰ IL ²⁶⁰ XXN FS-PPYISIKWSPKQIAAF ¹⁷⁰ L ¹⁸⁰ K ¹⁹⁰ S ²⁰⁰ Y ²¹⁰ SR ²²⁰ K ²³⁰ I ²⁴⁰ K ²⁵⁰ --GPK ²⁶⁰ IMAPE ²⁷⁰ CVHFVPEY ²⁸⁰ ND ²⁹⁰ AI ³⁰⁰ LL ³¹⁰ XXN
--	---

RKN, Rad

<i>R.sim_ABZ78968</i> <i>Mh_Contig2188.1417..3259</i> <i>Minc18739</i> <i>Minc11164</i> <i>Minc18650</i> <i>Minc18912</i> <i>Minc09178</i>	█ 230 █ 240 █ 250 █ 260 █ 270 DSSAVSQVSI IAGHIYGSGLADYSSAQNKQGVWMTEHYNAGFDWT SMMATAKEIH DKDAAAADVDI DAIHGYYGFFMSPQPVVKSGKPFWMTEHAI DGNKWNSVMQTAKDIH DPQSASAVDI IASHGYGFIMKPQPLVKKSGKKFWMTTEHFIDGNDFNSVMKQAEDIH DPQSAAAVDI IASHGYGFIMKPQPLVKKSGKQFWMTTEHCIDAGDFNSVMKI AEHIH DPQSAAAVDI IASHGYGFIMKSQPLVKKSGKQFWMTTEHCIDAGDFNSVMKI AEQIH DPQSAAAVDI DAMHGYGFIMKPQPSVIKSGKQFWMTTEHCIDAGDFNSVMKI AEQIH NPDVAKGVDDIAWHMYGMQLVSQTKAQMKGKSAWMTEKTNDGNDWKSFMETAKDIH
--	--

<i>R.sim_ABZ78968</i> <i>Mh_Contig2188.1417..3259</i> <i>Minc18739</i> <i>Minc11164</i> <i>Minc18650</i> <i>Minc18912</i> <i>Minc09178</i>	█ 290 █ 300 █ 310 █ 320 █ 330 DAMTVASYNAYVWWWFVDLN--NEFTSLTDKSGNPTKRGYIMAQWSKYIRPGYTRV DFMTIAEVTAAYVHHWLKSTNPSSQNMYLLYQNMQLTPKAFTVIGHFAKFIRPGYFRV NCLTIAEFNAYIHWWLRGN---SITMMLLYQNWQLTPKAYVIGHFAKFIRPGYFRV NSLTIAEFNAYVHHWLREPVP-SKMMLLFFQNKQLTPKAYVIGHFAKFIRPGYFRV NSLTIAEFNAYVHHWLREPVP-SKMMLLSFQDKQLTPKAYVIGHFAKFIRPGYVRV NSLTIAEFNAYIHWWLRDTSP-SKTMILLFQNKQLTPKAYVIGHFAKFIRPGYFRV DCMTIANYNAYVYFWFKDP---KYVSIVDNNYEITSRGYILGQYAKYIRPGYFRI
--	--



F) EXPN

M.jav_CAP59535

Minc10365

M.jav_CAP59536

Minc10366

Minc00158

M.jav_CAP59538

Minc18856

Minc10987

Minc10988

Minc07955

Minc07960

Minc14116

Mh10g200708_Contig790:11104..12567

G.rost_ACN58322

G.pal_scaf01091

Mh10g200708_Contig1277:27768..28838

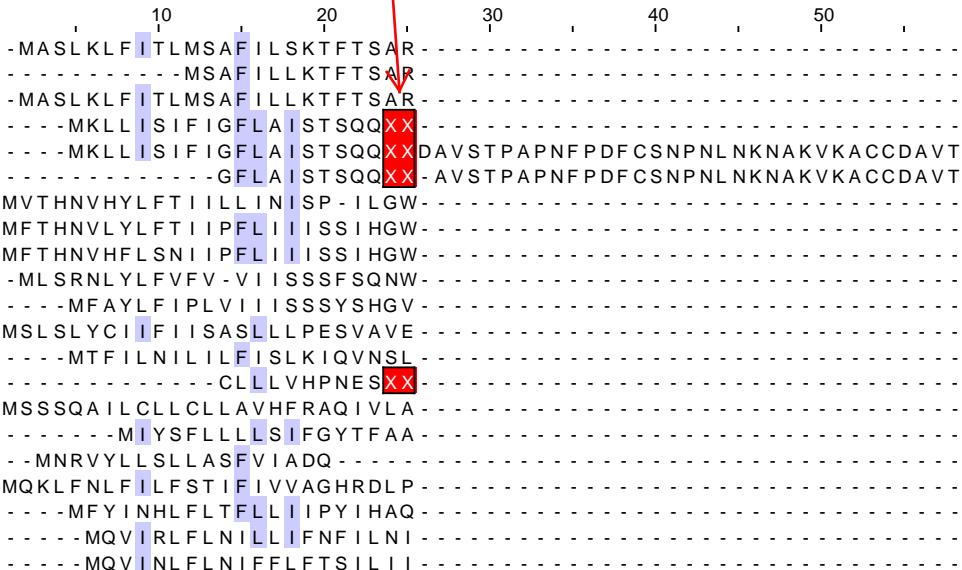
B.muc_BAG16533

Mh10g200708_Contig1905:1608..2586

Minc11468

Mh10g200708_Contig1905:330..998

Mh10g200708_Contig2646:6246..7094



M.jav_CAP59535

Minc10365

M.jav_CAP59536

Minc10366

Minc00158

M.jav_CAP59538

Minc18856

Minc10987

Minc10988

Minc07955

Minc07960

Minc14116

Mh10g200708_Contig790:11104..12567

G.rost_ACN58322

G.pal_scaf01091

Mh10g200708_Contig1277:27768..28838

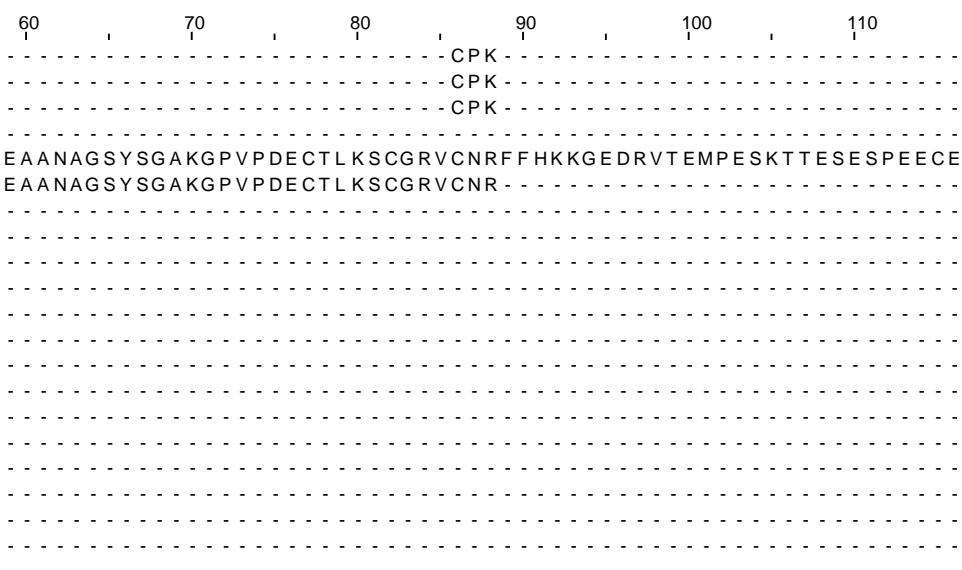
B.muc_BAG16533

Mh10g200708_Contig1905:1608..2586

Minc11468

Mh10g200708_Contig1905:330..998

Mh10g200708_Contig2646:6246..7094



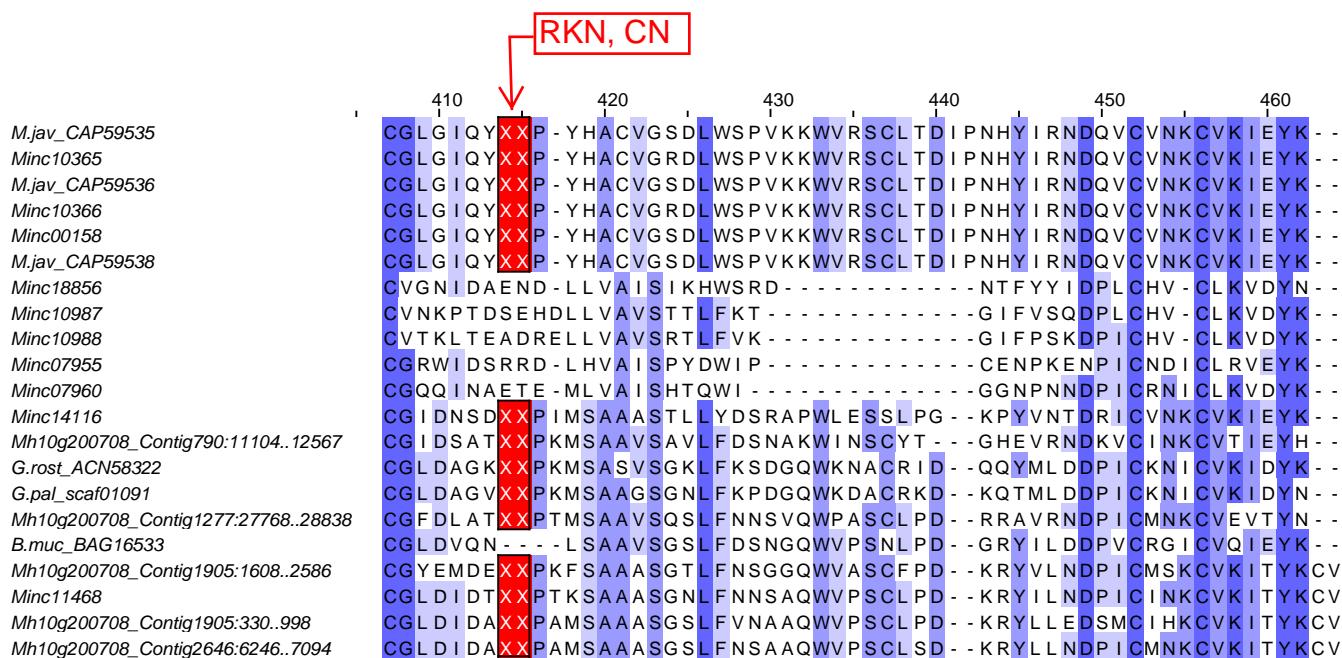
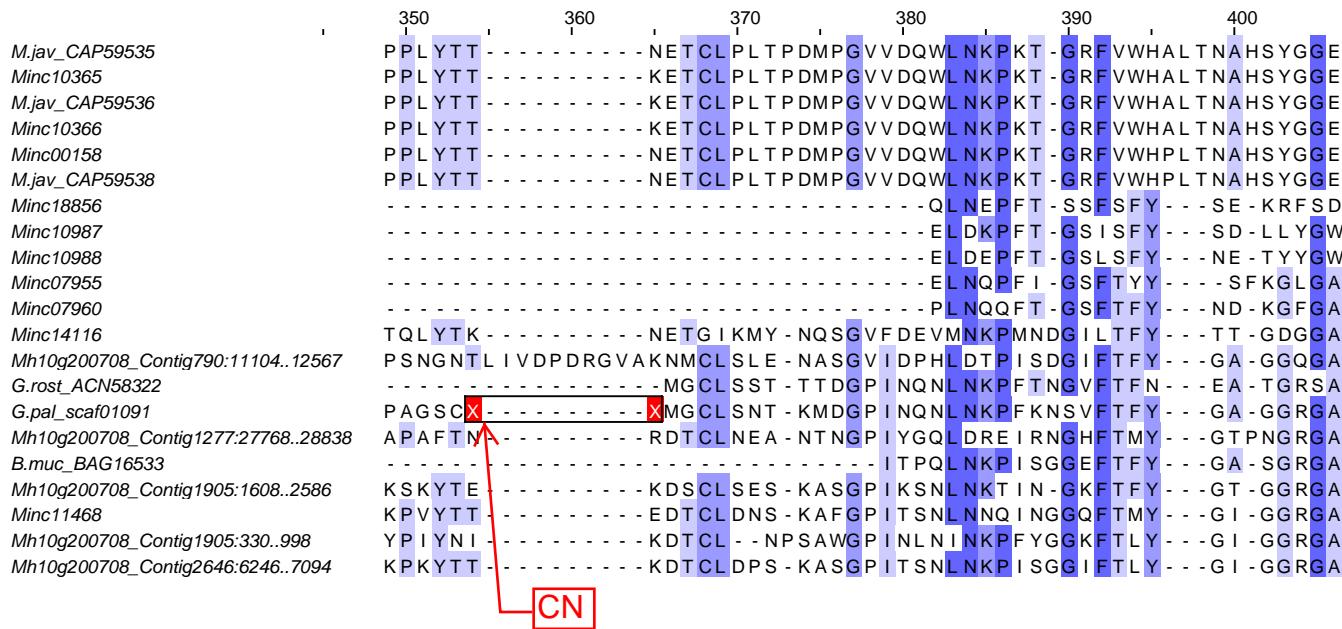
	120	130	140	150	160	170
<i>M.jav_CAP59535</i>	-	-	-	GTHPKNGEKSTINP	-INNPPAVG	T
<i>Minc10365</i>	-	-	-	GTHPKNGEKSTINP	-INNPPAVG	T
<i>M.jav_CAP59536</i>	-	-	-	GTHPKNGEKSTINP	-INNPPAVG	T
<i>Minc10366</i>	-	-	-	-	-APSTPQPGYPQPPA	APTT
<i>Minc00158</i>	CGDSGDETTGAQPTS	RNVYPSG	PESSTPGGQVASHK	PAPSTPQPGC	HQPPA	APTT
<i>M.jav_CAP59538</i>	-	-	-	SGPESSTPGGQVASHK	PAPSTPQPGC	HQPPA
<i>Minc18856</i>	-	-	-	-	-	-
<i>Minc10987</i>	-	-	-	-	-	-
<i>Minc10988</i>	-	-	-	-	-	-
<i>Minc07955</i>	-	-	-	-	-	-
<i>Minc07960</i>	-	-	-	-	-	-
<i>Minc14116</i>	-	-	-	KCTVSIDKNVLKSDNNNSHVQLIF	-	-
<i>Mh10g200708_Contig790:11104..12567</i>	-	-	-	PINSVISYNVLESRRGSSNIQIII	KNNGLTPICSVKF	-
<i>G.rost_ACN58322</i>	-	-	-	-	-	-
<i>G.pal_scaf01091</i>	-	-	-	-	TVTATLKAKTWNGGGQYTANFKNN	-
<i>Mh10g200708_Contig1277:27768..28838</i>	-	-	-	-	-	-
<i>B.muc_BAG16533</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:1608..2586</i>	-	-	-	-	-	-
<i>Minc11468</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:330..998</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig2646:6246..7094</i>	-	-	-	-	-	-

	180	190	200	210	220	230
<i>M.jav_CAP59535</i>	HPPVQP	-PSNKPPSPGYFS	-	-	-	-
<i>Minc10365</i>	HPPVQP	-PSNKPPSPGYFS	-	-	-	-
<i>M.jav_CAP59536</i>	HPPVQP	-PSNKPPSPGYFS	-	-	-	-
<i>Minc10366</i>	KPPRNPyPSGPEPTPKQPA	PSS	-	-	-	-
<i>Minc00158</i>	KPPRNLYPSGPEPTPKQPGY	PQGPASTPKQPGY	PQGPASTPKQPGY	PQGPASTPKQPGY	PQGRASPTPRNV	-
<i>M.jav_CAP59538</i>	KPPRNLYPSGPEPTPKQPGY	PQGPASTPKQPGY	PQGPASTPKQPGY	PQGPASTPKQPGY	PQGRASPTPRNV	-
<i>Minc18856</i>	-	-	-	-	-	-
<i>Minc10987</i>	-	-	-	-	-	-
<i>Minc10988</i>	-	-	-	-	-	-
<i>Minc07955</i>	-	-	-	-	-	-
<i>Minc07960</i>	-	-	-	-	-	-
<i>Minc14116</i>	KNMGQNPICFVKFSVLLPKG	XXCVTSNRNIKKDGNNYT	I	NRVIEPGNSYKDTCLGFN	-	-
<i>Mh10g200708_Contig790:11104..12567</i>	FLSLPNG	XXSVTQSNRSMRLRGNNKYRTPKNVR	I	E PGSSFDAGLI	I	I KGNKGKTLYL
<i>G.rost_ACN58322</i>	-	-	-	-	-	-
<i>G.pal_scaf01091</i>	DGSKIACSVKFSLTPKKG	XXTTIG	-	-	-	-
<i>Mh10g200708_Contig1277:27768..28838</i>	-	-	-	-	-	-
<i>B.muc_BAG16533</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:1608..2586</i>	NGPKKPNNPKQPTNP	NPPSFPI	-	-	-	-
<i>Minc11468</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:330..998</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig2646:6246..7094</i>	-	-	-	-	-	-

RKN
 CN

	240	250	260	270	280	
<i>M.jav_CAP59535</i>	-	-	PPSSNL PKQPSPPGK	-	-	YPSG
<i>Minc10365</i>	-	-	PPSSNL PKQPSPPGK	-	-	YPSG
<i>M.jav_CAP59536</i>	-	-	PPSSNL PKQPSPPGK	-	-	YPSG
<i>Minc10366</i>	-	-	PPSSNL PKQPSPPGK	-	-	YPSG
<i>Minc00158</i>	YPSGPEPSTPGGVASHKPPAAPSTPQPGYPQPPAAPT	KPPRNL	YPSGPEPTPKQPGY	-	-	-
<i>M.jav_CAP59538</i>	YPSGPEPSTPGGVASHKPPAAPSTPQPGYPQPPAAPT	KPPRNL	YPSGPEPTPKQPGY	-	-	-
<i>Minc18856</i>	-	-	-	-	-	-
<i>Minc10987</i>	-	-	-	-	-	-
<i>Minc10988</i>	-	-	-	-	-	-
<i>Minc07955</i>	-	-	-	-	-	-
<i>Minc07960</i>	-	-	-	-	-	-
<i>Minc14116</i>	GTGIPNFKVISTKKCGTPKPKE	LTTLKSPKE	IPSTNKPTLPSQT	SNGIPTEPTQELNV	NEDNDEKDCTTT	E
<i>Mh10g200708_Contig790:11104..12567</i>	NVTECNNK	-	-	-	-	-
<i>G.rost_ACN58322</i>	-	-	-	-	-	-
<i>G.pal_scaf01091</i>	-	-	-	-	NVW	-
<i>Mh10g200708_Contig1277:27768..28838</i>	-	-	-	-	-	-
<i>B.muc_BAG16533</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:1608..2586</i>	-	-	-	-	-	-
<i>Minc11468</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:330..998</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig2646:6246..7094</i>	-	-	-	-	-	-

	300	310	320	330	340	
<i>M.jav_CAP59535</i>	PPNHNQPKQPSPTPPSSGPSPSSAPSVPAL	PSGQYPPGSN	-	VTSNSLT	TV	
<i>Minc10365</i>	PPSHNQPKQPS	-	PSPPSSAPSVPAPP	SGQYPPGSN	VTSNSLT	TV
<i>M.jav_CAP59536</i>	PPSHNQPKQPS	-	PSPPSSAPSVPAPP	SGQYPPGSN	VTSNSLT	TV
<i>Minc10366</i>	-	-	PTPPSSAPSVPAPP	SGQYPPGSN	VTSNSLT	TV
<i>Minc00158</i>	PQGPASPTPPSHNQPKQPSPTPPSSAPSPPGL	PPSVTAPP	SGQYPPGSN	VTSNSLT	TV	
<i>M.jav_CAP59538</i>	PQGPASPTPPSHKQPKQPSPTPPSSAPSPPGL	PPSVTAPP	SGQYPPGSN	VTSNSLT	TV	
<i>Minc18856</i>	-	-	-	-	-	-
<i>Minc10987</i>	-	-	-	-	-	-
<i>Minc10988</i>	-	-	-	-	-	-
<i>Minc07955</i>	-	-	-	-	-	-
<i>Minc07960</i>	-	-	-	-	-	-
<i>Minc14116</i>	TPTPSSLNNNTNNQPPQPTPNNNNI	LPKPLRHQGELPFESVINYFTGINKIGGDQMVI	-	-	-	-
<i>Mh10g200708_Contig790:11104..12567</i>	SSGKNKGNEKIKTTTESNENNDK	ITKSNKGNNKEITTTTESSGNNGNN	-	-	-	-
<i>G.rost_ACN58322</i>	-	-	-	-	-	-
<i>G.pal_scaf01091</i>	NAVSGASNQYTLAPPADFGPGATHNAGVNINGNGAP	TLKLIEAKYFIN	X	DVC	GGAA	
<i>Mh10g200708_Contig1277:27768..28838</i>	-	-	-	-	-	-
<i>B.muc_BAG16533</i>	-	-	-	-	-	-
<i>Mh10g200708_Contig1905:1608..2586</i>	-	-	KPNPPKKPSNPKPPNP	I	KPIHPI	-NDGNTKVV
<i>Minc11468</i>	-	-	-	-	-	-DSLEV
<i>Mh10g200708_Contig1905:330..998</i>	-	-	-	-	-	-HAQNDLTV
<i>Mh10g200708_Contig2646:6246..7094</i>	-	-	-	-	-	-KAQNSLVV



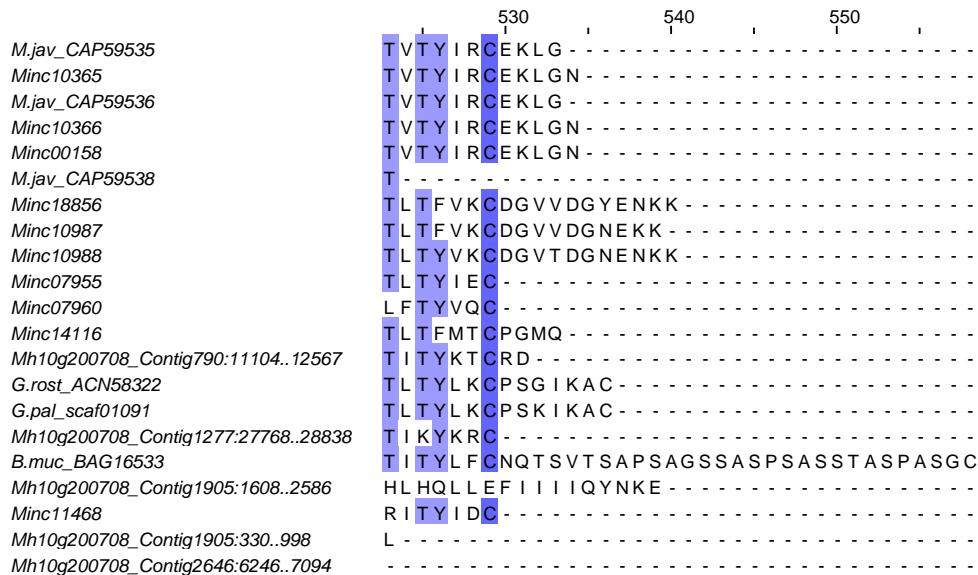
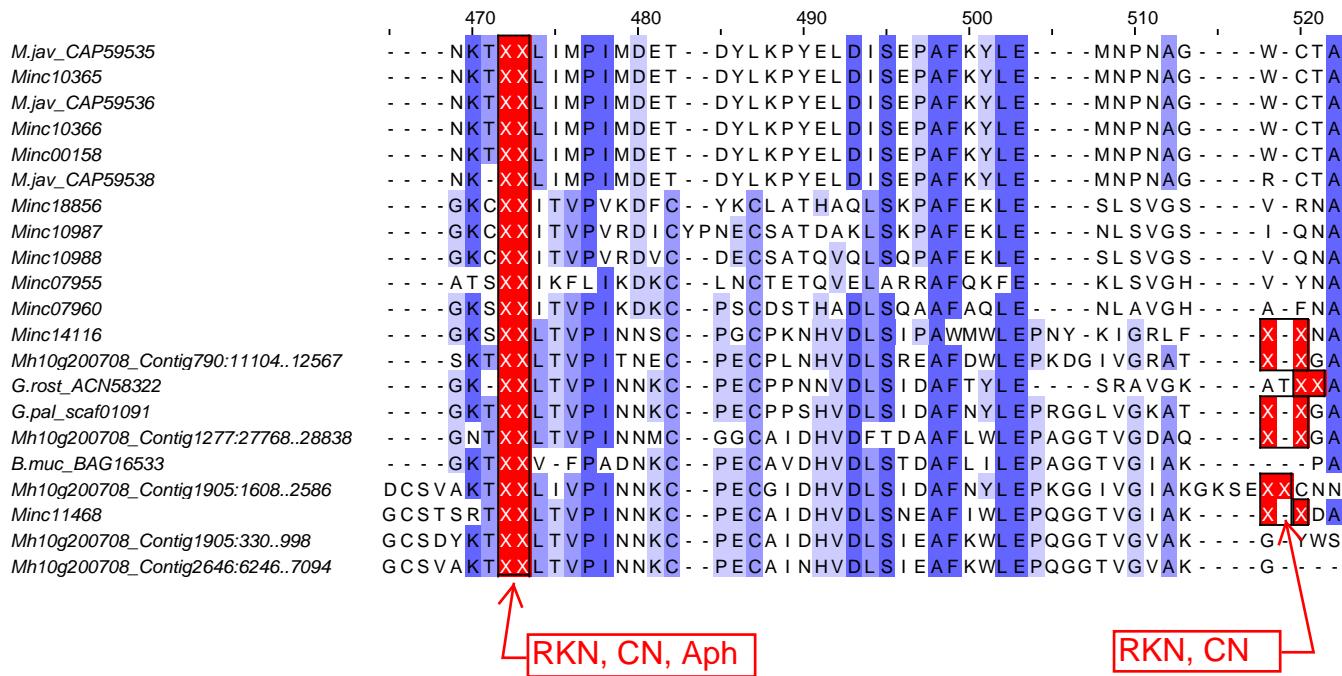


Table S3

GC percent	Minc CDS	PCWM CDS	GC percent	Minc CDS	PCWM CDS
16	2	0	mean	36,26	36,33
17	3	0	median	36,00	36,50
18	4	0			
19	6	0			
20	14	0			
21	28	0			
22	56	0			
23	103	0			
24	144	0			
25	237	0			
26	305	1			
27	376	0			
28	452	1			
29	615	0			
30	764	1			
31	948	5			
32	1109	5			
33	1270	4			
34	1404	8			
35	1538	10			
36	1573	7			
37	1514	9			
38	1484	9			
39	1222	7			
40	998	6			
41	922	9			
42	747	3			
43	525	0			
44	381	0			
45	303	0			
46	261	0			
47	222	0			
48	154	0			
49	122	0			
50	88	0			
51	95	0			
52	95	0			
53	65	0			
54	51	0			
55	41	0			
56	30	0			
57	30	0			
58	16	0			
59	9	0			
60	18	0			
61	8	0			
62	6	0			
63	1	0			
64	3	0			
65	3	0			

Fig. S12

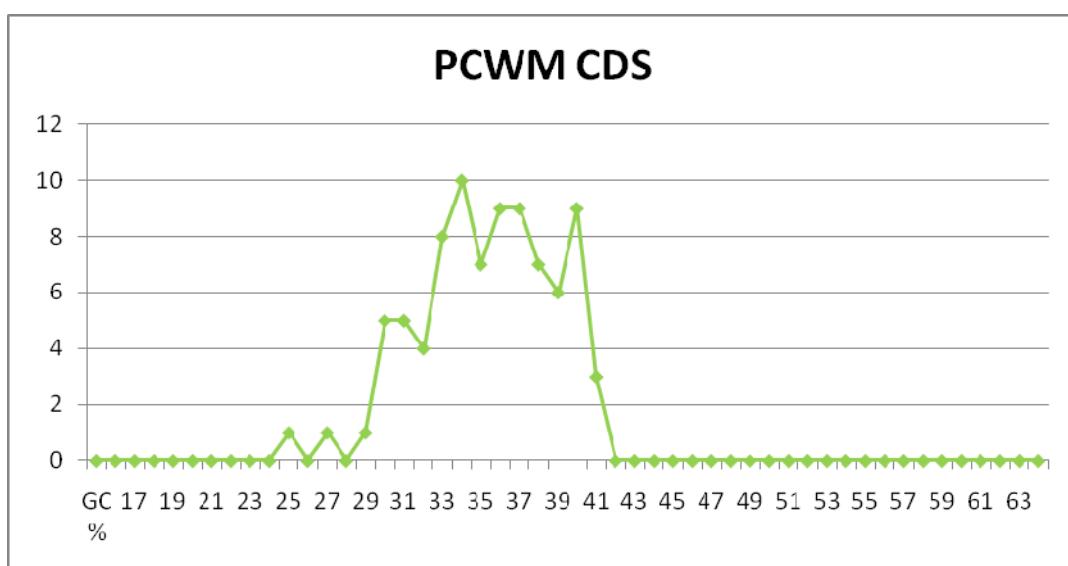
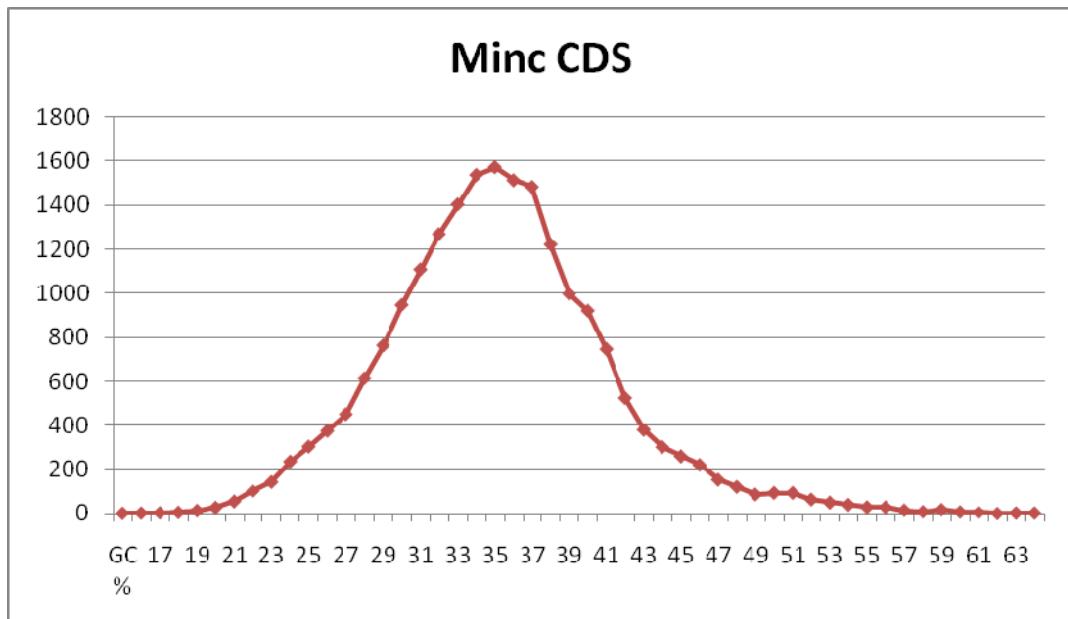


Table S4A

Codon	aa	PCWM CDS			All CDS			% diff.
		Fraction	Frequency	Number	Fraction	Frequency	Number	
GCA	A	0,359	22,068	607	0,346	18,107	130744	1,3
GCC	A	0,119	7,307	201	0,127	6,614	47755	0,8
GCG	A	0,073	4,508	124	0,082	4,262	30772	0,9
GCT	A	0,449	27,667	761	0,446	23,295	168209	0,3
TGC	C	0,278	7,453	205	0,274	5,665	40902	0,4
TGT	C	0,722	19,341	532	0,726	15,037	108574	0,4
GAC	D	0,254	12,543	345	0,248	11,978	86490	0,6
GAT	D	0,746	36,792	1012	0,752	36,417	262955	0,6
GAA	E	0,722	26,831	738	0,768	55,539	401029	4,6
GAG	E	0,278	10,325	284	0,232	16,748	120932	4,6
TTC	F	0,238	10,507	289	0,211	9,879	71332	2,7
TTT	F	0,762	33,629	925	0,789	36,997	267146	2,7
GGA	G	0,478	39,228	1079	0,467	25,9	187013	1,1
GGC	G	0,164	13,415	369	0,139	7,719	55737	2,5
GGG	G	0,135	11,052	304	0,137	7,586	54774	0,2
GGT	G	0,223	18,323	504	0,256	14,204	102559	3,3
CAC	H	0,285	4,472	123	0,254	5,396	38963	3,1
CAT	H	0,715	11,234	309	0,746	15,851	114456	3,1
ATA	I	0,309	22,431	617	0,258	17,215	124303	5,1
ATC	I	0,119	8,653	238	0,107	7,105	51305	1,2
ATT	I	0,571	41,446	1140	0,635	42,328	305637	6,4
AAA	K	0,797	55,442	1525	0,774	61,213	442002	2,3
AAG	K	0,203	14,142	389	0,226	17,874	129061	2,3
CTA	L	0,084	5,344	147	0,073	6,6	47658	1,1
CTC	L	0,069	4,363	120	0,061	5,52	39861	0,8
CTG	L	0,055	3,454	95	0,047	4,231	30551	0,8
CTT	L	0,253	15,997	440	0,237	21,283	153681	1,6
TTA	L	0,344	21,813	600	0,335	30,136	217601	0,9
TTG	L	0,195	12,361	340	0,246	22,133	159815	5,1
ATG	M	1	17,96	494	1	21,04	151925	0
AAC	N	0,253	21,377	588	0,202	12,701	91710	5,1
AAT	N	0,747	63,15	1737	0,798	50,14	362044	5,1
CCA	P	0,415	21,704	597	0,414	19,22	138783	0,1
CCC	P	0,094	4,908	135	0,096	4,459	32194	0,2
CCG	P	0,064	3,345	92	0,101	4,702	33950	3,7
CCT	P	0,427	22,359	615	0,389	18,061	130409	3,8
CAA	Q	0,814	29,412	809	0,808	37,116	267999	0,6
CAG	Q	0,186	6,726	185	0,192	8,796	63514	0,6
AGA	R	0,353	9,961	274	0,33	16,597	119842	2,3
AGG	R	0,093	2,618	72	0,108	5,436	39249	1,5
CGA	R	0,169	4,763	131	0,169	8,497	61354	0
CGC	R	0,082	2,327	64	0,079	3,975	28699	0,3
CGG	R	0,031	0,873	24	0,05	2,522	18211	1,9
CGT	R	0,273	7,707	212	0,265	13,326	96222	0,8
AGC	S	0,095	7,671	211	0,084	6,331	45716	1,1
AGT	S	0,207	16,724	460	0,188	14,124	101987	1,9
TCA	S	0,284	22,904	630	0,264	19,8	142969	2
TCC	S	0,065	5,272	145	0,079	5,931	42828	1,4
TCG	S	0,079	6,399	176	0,074	5,533	39949	0,5
TCT	S	0,268	21,632	595	0,311	23,359	168665	4,3
ACA	T	0,427	26,176	720	0,436	22,539	162747	0,9
ACC	T	0,076	4,654	128	0,096	4,942	35687	2
ACG	T	0,073	4,472	123	0,087	4,489	32410	1,4
ACT	T	0,425	26,067	717	0,382	19,738	142525	4,3
GTA	V	0,249	14,615	402	0,199	9,988	72119	5
GTC	V	0,109	6,362	175	0,12	6,009	43391	1,1
GTG	V	0,109	6,362	175	0,133	6,656	48059	2,4
GTT	V	0,533	31,266	860	0,548	27,428	198047	1,5
TGG	W	1	15,887	437	1	10,682	77132	0
TAC	Y	0,204	7,998	220	0,222	6,724	48550	1,8
TAT	Y	0,796	31,266	860	0,778	23,575	170229	1,8
TAA	*	0,395	1,163	32	0,538	1,472	10628	14,3
TAG	*	0,099	0,291	8	0,148	0,405	2921	4,9
TGA	*	0,506	1,491	41	0,314	0,858	6193	19,2

Table S4B

Codon		Usage in PCWD (%)	Usage in the whole genome (%)	% difference
TAA	STOP	39.5	53.8	14.3
TAG		9.9	14.8	4.9
TGA		50.6	31.4	19.2
AAC	ASN	25.3	20.2	5.1
AAT		74.7	79.8	5.1
ATA	ILE	30.9	25.8	5.1
ATC		11.9	10.7	1.2
ATT		57.1	63.5	6.4
CTA	LEU	8.4	7.3	1.1
CTC		6.9	6.1	0.8
CTG		5.5	4.7	0.8
CTT		25.3	23.7	1.6
TTA		34.4	33.5	0.9
TTG		19.5	24.6	5.1

While in the whole genome the most used stop codon is TAA (53.8%) this codon is only used 39.5 % in PCWD genes and in this dataset the most used stop codon is TGA (50.6 %). Other codons where a No other substantial differences in codon usage were observed between PCWM genes and the rest of *M. incognita* genes (see table S4A).

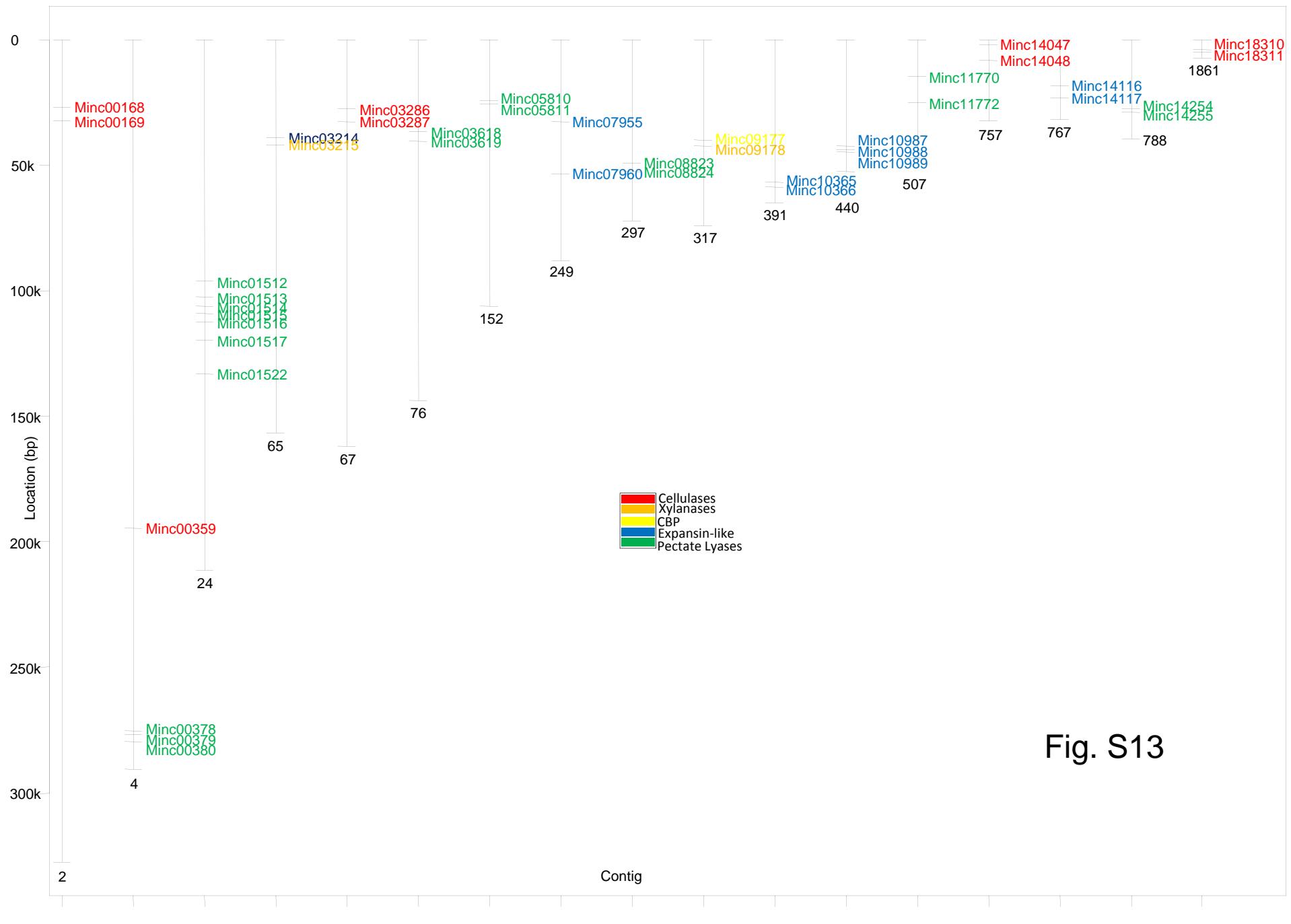


Fig. S13

Fig. S14

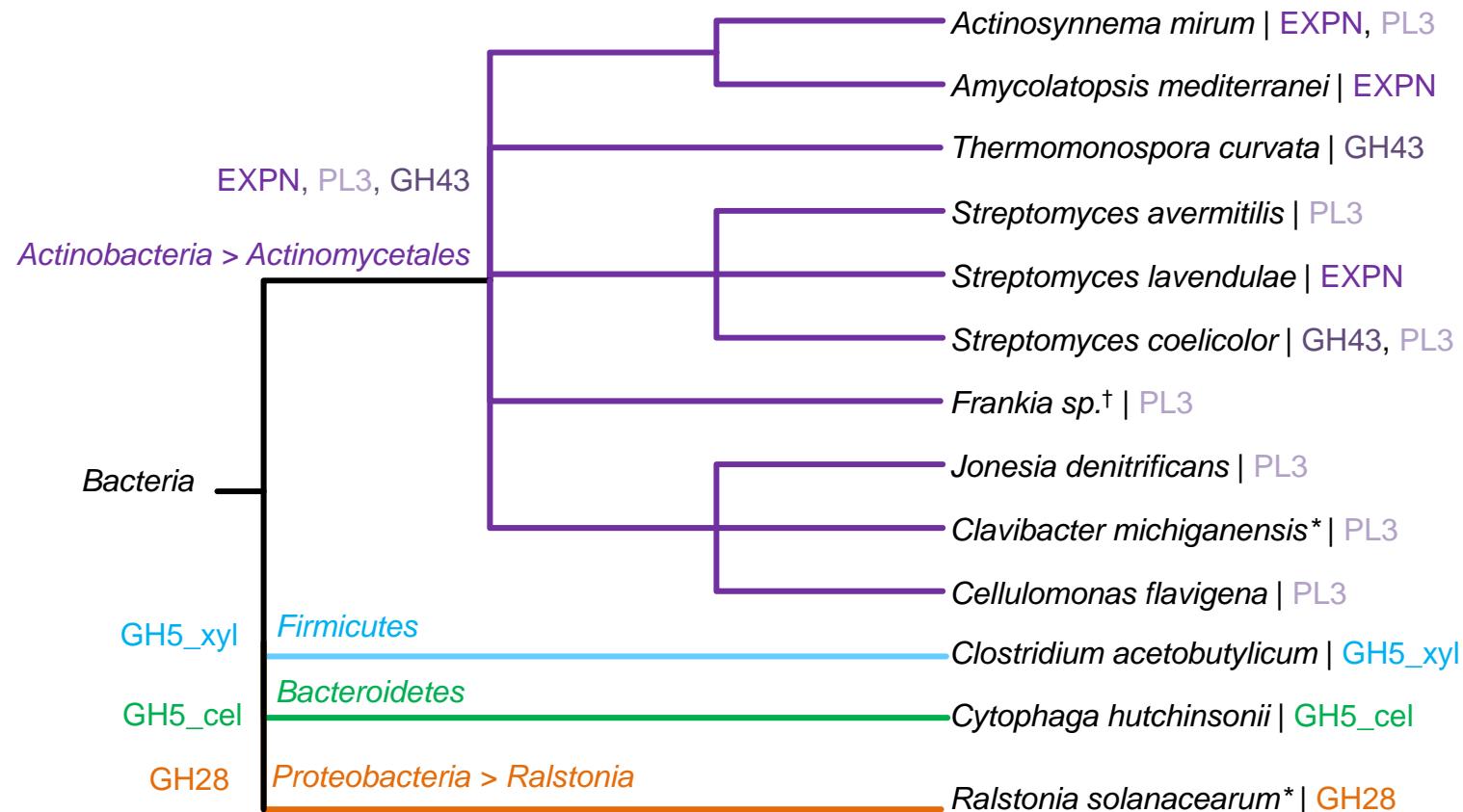


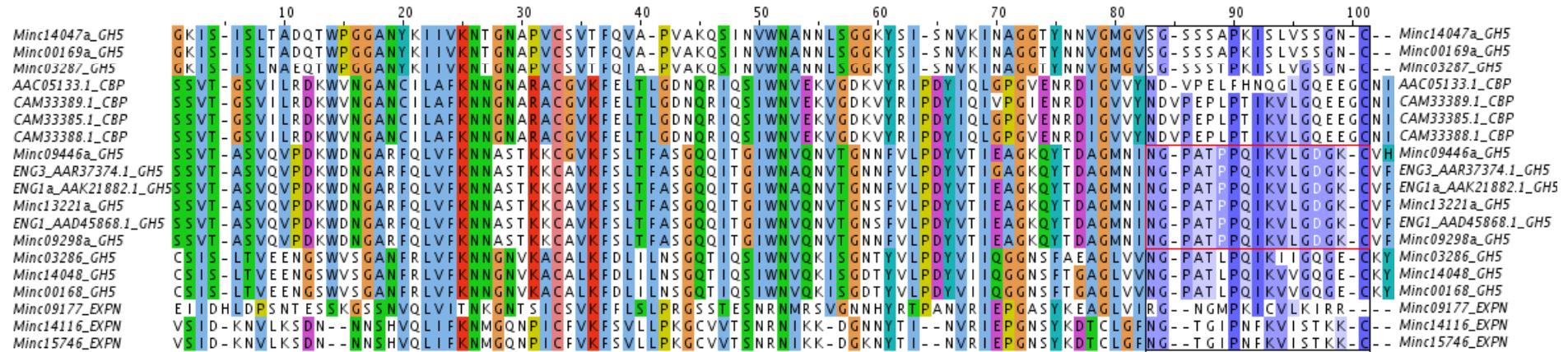
Table S5. Statistical characteristics, evolutionary models and parameters of phylogenetic analyses.

Family	Fig.	alpha	Pi	Model (PP [†])	# Generation / # Bootstrap	Method*
GH28	1A, S2A, S2B	1.886	0.013	WAG (1.0)	200,000	B
GH28	1A, S2A, S2B	1.745	0.008	WAG (NA)	100	ML
PL3	S3A, S3C	1.148	0.003	WAG (1.0)	180,000	B
PL3	S3A	1.073	0.001	WAG (NA)	100	ML
PL3	1B, S3B	1.613	0.032	WAG (1.0)	350,000	B
PL3	1B, S3B	1.465	0.018	WAG (NA)	100	ML
GH43	1C, S4A	1.395	0.054	WAG (1.0)	150,000	B
GH43	1C, S4A	1.226	0.033	WAG (NA)	100	ML
GH5_cel	S5A, S5C	1.057	0.032	WAG (1.0)	600,000	B
GH5_cel	S5A	0.875	0.016	WAG (NA)	100	ML
GH5_cel	2A, S5B	1.339	0.042	WAG (1.0)	350,000	B
GH5_cel	2A, S5B	1.196	0.027	WAG (NA)	100	ML
GH5_xyl	S6A	1.929	0.008	WAG (1.0)	100,000	B
GH5_xyl	S6A	1.849	0.004	WAG (NA)	100	ML
GH5_xyl	2B, S6B	2.015	0.017	WAG (1.0)	400,000	B
GH5_xyl	2B, S6B	1.789	0.011	WAG (NA)	100	ML
EXPN	S7A	1.589	0.021	WAG (1.0)	800,000	B
EXPN	S7A	1.513	0.006	WAG (NA)	100	ML
EXPN	2C, S7B	1.901	0.021	WAG (1.0)	500,000	B
EXPN	2C, S7B	1.968	0.006	WAG (NA)	100	ML
CBM2	S8A	6.643	0.025	WAG (1.0)	200,000	B
CBM2	S8A	5.971	0.008	WAG (NA)	100	ML
CBM2	S8B	22.697	0.023	WAG (1.0)	150,000	B
CBM2	S8B	7.849	0.009	WAG (NA)	100	ML

* B stands for Bayesian approach and ML stands for Maximum Likelihood.

[†] Posterior Probability for the best evolutionary models. By definition, this only applies for Bayesian phylogenies and the best model was chosen for Maximum Likelihood phylogenies. NA thus stands for “non applicable”.

Fig. S15



SI References

1. Abad P, et al. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat Biotechnol* 26(8):909-915
2. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402
3. Cantarel BL, et al. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37(Database issue):D233-238
4. Opperman CH, et al. (2008) Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc Natl Acad Sci U S A* 105(39):14802-14807
5. Martin J, et al. (2009) Nematode.net update 2008: improvements enabling more efficient data mining and comparative nematode genomics. *Nucleic Acids Res* 37(Database issue):D571-578
6. Wasmuth J, Schmid R, Hedley A, & Blaxter M (2008) On the extent and origins of genic novelty in the phylum nematoda. *PLoS neglected tropical diseases* 2(7):e258
7. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113
8. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797
9. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*
10. Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754-755
11. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690
12. Shimodaira H & Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17(12):1246-1247
13. Rice P, Longden I, & Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276-277
14. Broman KW, Wu H, Sen S, & Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7):889-890
15. Keller O, Odronitz F, Stanke M, Kollmar M, & Waack S (2008) Scipio: using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. *BMC Bioinformatics* 9:278
16. Odronitz F, Pillmann H, Keller O, Waack S, & Kollmar M (2008) WebScipio: an online tool for the determination of gene structures using protein sequences. *BMC Genomics* 9:422
17. Pollastri G, Martin AJ, Mooney C, & Vullo A (2007) Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information. *BMC Bioinformatics* 8:201
18. de Almeida Engler J, et al. (2004) Dynamic cytoskeleton rearrangements in giant cells and syncytia of nematode-infected roots. *Plant J* 38(1):12-26
19. Jaubert S, et al. (2005) In planta secretion of a calreticulin by migratory and sedentary stages of root-knot nematode. *Mol Plant Microbe Interact* 18(12):1277-1284
20. Holterman M, et al. (2006) Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Mol Biol Evol* 23(9):1792-1800
21. van Meegen H, et al. (2009) A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11:927-950
22. Blaxter ML, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392(6671):71-75
23. Holterman M, et al. (2009) Small subunit rDNA-based phylogeny of the Tylenchida sheds light on relationships among some high-impact plant-parasitic nematodes and the evolution of plant feeding. *Phytopathology* 99(3):227-235
24. Muse BD, Moore LD, Muse RR, & Williams AS (1970) Pectolytic and Cellulolytic Enzymes of Two Populations of *Ditylenchus dipsaci* on 'Wando' Pea (*Pisum sativum L.*). *J Nematol* 2(2):118-124
25. Chitwood DJ & Krusberg LR (1977) Pectolytic Enzymes in Three Populations of *Ditylenchus dipsaci*. *J Nematol* 9(3):187-192
26. Bird DM, et al. (2009) The genomes of root-knot nematodes. *Annu Rev Phytopathol* 47:333-351
27. Jaubert S, Laffaire JB, Abad P, & Rosso MN (2002) A polygalacturonase of animal origin isolated from the root-knot nematode *Meloidogyne incognita*. *FEBS Lett* 522(1-3):109-112
28. Rosso MN, Dubrana MP, Cimbolini N, Jaubert S, & Abad P (2005) Application of RNA interference to root-knot nematode genes encoding esophageal gland proteins. *Mol Plant Microbe Interact* 18(7):615-620
29. Karim N, Jones JT, Okada H, & Kikuchi T (2009) Analysis of expressed sequence tags and identification of genes encoding cell-wall-degrading enzymes from the fungivorous nematode *Aphelenchus avenae*. *BMC Genomics* 10(1):525

30. Kikuchi T, Shibuya H, Aikawa T, & Jones JT (2006) Cloning and characterization of pectate lyases expressed in the esophageal gland of the pine wood nematode *Bursaphelenchus xylophilus*. *Mol Plant Microbe Interact* 19(3):280-287
31. Popeijus H, et al. (2000) Degradation of plant cell walls by a nematode. *Nature* 406(6791):36-37
32. Gao B, et al. (2003) The parasitome of the phytonematode *Heterodera glycines*. *Mol Plant Microbe Interact* 16(8):720-726
33. De Boer JM, et al. (2002) Cloning of a Putative Pectate Lyase Gene Expressed in the Subventral Esophageal Glands of *Heterodera glycines*. *J Nematol* 34(1):9-11
34. Vanholme B, et al. (2007) Molecular characterization and functional importance of pectate lyase secreted by the cyst nematode *Heterodera schachtii*. *Molecular Plant Pathology* 8(3):267-278
35. Huang G, et al. (2005) Developmental expression and molecular analysis of two *Meloidogyne incognita* pectate lyase genes. *Int J Parasitol* 35(6):685-692
36. Doyle EA & Lambert KN (2002) Cloning and characterization of an esophageal-gland-specific pectate lyase from the root-knot nematode *Meloidogyne javanica*. *Mol Plant Microbe Interact* 15(6):549-556
37. Wei YD, et al. (2006) Molecular cloning, expression, and enzymatic activity of a novel endogenous cellulase from the mulberry longicorn beetle, *Apriona germari*. *Comp Biochem Physiol B Biochem Mol Biol* 145(2):220-229
38. Sugimura M, Watanabe H, Lo N, & Saito H (2003) Purification, characterization, cDNA cloning and nucleotide sequencing of a cellulase from the yellow-spotted longicorn beetle, *Psacothea hilaris*. *Eur J Biochem* 270(16):3455-3460
39. Kyndt T, Haegeman A, & Gheysen G (2008) Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. *BMC Evol Biol* 8(1):305
40. Smant G, et al. (1998) Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci U S A* 95(9):4906-4911
41. Rehman S, et al. (2009) Identification and characterization of the most abundant cellulases in stylet secretions from *Globodera rostochiensis*. *Phytopathology* 99(2):194-202
42. Yan Y, et al. (1998) Genomic organization of four beta-1,4-endoglucanase genes in plant-parasitic cyst nematodes and its evolutionary implications. *Gene* 220(1-2):61-70
43. Goellner M, Smant G, De Boer JM, Baum TJ, & Davis EL (2000) Isolation of Beta-1,4-Endoglucanase Genes from *Globodera tabacum* and their Expression During Parasitism. *J Nematol* 32(2):154-165
44. Wang X, et al. (1999) In planta localization of a beta-1,4-endoglucanase secreted by *Heterodera glycines*. *Mol Plant Microbe Interact* 12(1):64-67
45. de Boer JM, et al. (1999) Developmental expression of secretory beta-1,4-endoglucanases in the subventral esophageal glands of *Heterodera glycines*. *Mol Plant Microbe Interact* 12(8):663-669
46. Yan Y, Smart G, & Davis E (2001) Functional screening yields a new beta-1,4-endoglucanase gene from *Heterodera glycines* that may be the product of recent gene duplication. *Mol Plant Microbe Interact* 14(1):63-71
47. Gao B, et al. (2002) Identification of a New ss-1,4-endoglucanase Gene Expressed in the Esophageal Subventral Gland Cells of *Heterodera glycines*. *J Nematol* 34(1):12-15
48. Vanholme B, et al. (2006) Detection of putative secreted proteins in the plant-parasitic nematode *Heterodera schachtii*. *Parasitology research* 98(5):414-424
49. De Meutter J, et al. (2001) Preparation and sequencing of secreted proteins from the pharyngeal glands of the plant parasitic nematode *Heterodera schachtii*. *Molecular Plant Pathology* 2(5):297-301
50. Ledger TN, Jaubert S, Bosselut N, Abad P, & Rosso MN (2006) Characterization of a new beta-1,4-endoglucanase gene from the root-knot nematode *Meloidogyne incognita* and evolutionary scheme for phytonematode family 5 glycosyl hydrolases. *Gene* 382:121-128
51. Bera-Maillet C, Arthaud L, Abad P, & Rosso MN (2000) Biochemical characterization of MI-ENG1, a family 5 endoglucanase secreted by the root-knot nematode *Meloidogyne incognita*. *Eur J Biochem* 267(11):3255-3263
52. Rosso MN, et al. (1999) Isolation of a cDNA encoding a beta-1,4-endoglucanase in the root-knot nematode *Meloidogyne incognita* and expression analysis during plant parasitism. *Mol Plant Microbe Interact* 12(7):585-591
53. Uehara T, Kushida A, & Momota Y (2001) PCR-based cloning of two beta-1,4-endoglucanases from the root-lesion nematode *Pratylenchus penetrans*. *Nematology* 3:335-341
54. Haegeman A, Jacob J, Vanholme B, Kyndt T, & Gheysen G (2008) A family of GHF5 endo-1,4-beta-glucanases in the migratory plant-parasitic nematode *Radopholus similis*. *Plant Pathology* 57(3):581-590
55. Mitreva-Dautova M, et al. (2006) A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode *Meloidogyne incognita*. *Mol Plant Microbe Interact* 19(5):521-529
56. Jacob J, Mitreva M, Vanholme B, & Gheysen G (2008) Exploring the transcriptome of the burrowing

- nematode *Radopholus similis*. *Mol Genet Genomics* 280(1):1-17
- 57. Haegeman A, Vanholme B, & Gheysen G (2009) Characterization of a putative endoxylanase in the migratory plant-parasitic nematode *Radopholus similis*. *Mol Plant Pathol* 10(3):389-401
 - 58. Haegeman A, et al. (2009) Expressed sequence tags of the peanut pod nematode *Ditylenchus africanus*: the first transcriptome analysis of an Anguinid nematode. *Mol Biochem Parasitol* 167(1):32-40
 - 59. Qin L, et al. (2004) Plant degradation: a nematode expansin acting on plants. *Nature* 427(6969):30
 - 60. Kudla U, et al. (2005) Origin, distribution and 3D-modeling of Gr-EXPB1, an expansin from the potato cyst nematode *Globodera rostochiensis*. *FEBS Lett* 579(11):2451-2457

Dataset S01. List of sequences and species included in the GH28 polygalacturonases phylogenetic analysis.

Accession numbers, source database, usual name (where applicable), abbreviated and full species name as well as taxonomic group are indicated for every sequence retained in the phylogenetic analysis of GH28 polygalacturonases. Sequences are sorted by taxonomic group then by species name. Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the _RKN suffix for root-knot nematode; Bact: Bacteria, with the following suffixes: _Spir: Spirochaetes, _Prot: Proteobacteria, _Fus: Fusobacteria, _Firm: Firmicutes, _Chla: Chlamydiae / Verrucomicrobia, _Bac: Bacteroidetes, _Acid: Acidobacteria; Plant: Plant; Arch: Archaea; Unk: unidentified organism.

Accession	Source	Usual name	Species	Species long	Group
Mh_Contig2637.1751..5651	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig1443.9536..11966	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig1886.4771..7593	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
AAM28240	NCBI	Mi-PG1	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc18543b	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc18899b	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAC63679	NCBI		A.thal	<i>Arabidopsis thaliana</i>	Plant
AAM14020	NCBI		A.thal	<i>Arabidopsis thaliana</i>	Plant
AAT85725	NCBI		A.thal	<i>Arabidopsis thaliana</i>	Plant
CAW63366	NCBI		G.max	<i>Glycine max</i>	Plant
EEE78584	NCBI		P.trich	<i>Populus trichocarpa</i>	Plant
EEE80827	NCBI		P.trich	<i>Populus trichocarpa</i>	Plant
EEE84316	NCBI		P.trich	<i>Populus trichocarpa</i>	Plant
EEE99249	NCBI		P.trich	<i>Populus trichocarpa</i>	Plant
CAW63324	NCBI		Z.may	<i>Zea mays</i>	Plant
AAT11785	NCBI		T.pect	<i>Treponema pectinovorum</i>	Bact_Spir
AAA24842	NCBI		E.chry	<i>Erwinia chrysanthemi</i>	Bact_Prot
CAB99318	NCBI		E.chry	<i>Erwinia chrysanthemi</i>	Bact_Prot
CAB99319	NCBI		E.chry	<i>Erwinia chrysanthemi</i>	Bact_Prot
CAB99320	NCBI		E.chry	<i>Erwinia chrysanthemi</i>	Bact_Prot
ABP60790	NCBI		E.sp	<i>Enterobacter sp.</i>	Bact_Prot
AAL49975	NCBI		K.oxy	<i>Klebsiella oxytoca</i>	Bact_Prot
CAG74100	NCBI		P.atro	<i>Pectobacterium atrosepticum</i>	Bact_Prot
CAG76010	NCBI		P.atro	<i>Pectobacterium atrosepticum</i>	Bact_Prot
ABQ78032	NCBI		P.puti	<i>Pseudomonas putida</i>	Bact_Prot
AAC46001	NCBI		R.sol	<i>Ralstonia solanacearum</i>	Bact_Prot
CAD15458	NCBI		R.sol	<i>Ralstonia solanacearum</i>	Bact_Prot
CAQ35716	NCBI		R.sol	<i>Ralstonia solanacearum</i>	Bact_Prot

CAQ61467	NCBI	R.sol	<i>Ralstonia solanacearum</i>	Bact_Prot
ABD80215	NCBI	S.deg	<i>Saccharophagus degradans</i>	Bact_Prot
ACR12997	NCBI	T.turne	<i>Teredinibacter turnerae</i>	Bact_Prot
AAC15064	NCBI	Y.ent	<i>Yersinia enterocolitica</i>	Bact_Prot
CAL10303	NCBI	Y.ent	<i>Yersinia enterocolitica</i>	Bact_Prot
ABD84266	NCBI	Y.sp	<i>Yersinia sp.</i>	Bact_Prot
EEP37217	NCBI	S.term	<i>Sebaldella termitidis</i>	Bact_Fus
ACV59463	NCBI	A.acido	<i>Alicyclobacillus acidocaldarius</i>	Bact_Firm
ABR47121	NCBI	A.meta	<i>Alkaliphilus metallireducens</i>	Bact_Firm
ACM60667	NCBI	A.ther	<i>Anaerocellum thermophilum</i>	Bact_Firm
ACM61449	NCBI	A.ther	<i>Anaerocellum thermophilum</i>	Bact_Firm
AAN26178	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
AAU42338	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
ABV63637	NCBI	B.pum	<i>Bacillus pumilus</i>	Bact_Firm
BAB85762	NCBI	Baci.sp	<i>Bacillus sp.</i>	Bact_Firm
AAK78335	NCBI	C.aceto	<i>Clostridium acetobutylicum</i>	Bact_Firm
AAK81605	NCBI	C.aceto	<i>Clostridium acetobutylicum</i>	Bact_Firm
ABX42928	NCBI	C.phyt	<i>Clostridium phytofermentans</i>	Bact_Firm
ABX43097	NCBI	C.phyt	<i>Clostridium phytofermentans</i>	Bact_Firm
ABP66002	NCBI	C.sacc	<i>Caldicellulosiruptor saccharolyticus</i>	Bact_Firm
ABP66284	NCBI	C.sacc	<i>Caldicellulosiruptor saccharolyticus</i>	Bact_Firm
ACR71465	NCBI	E.elig	<i>Eubacterium eligens</i>	Bact_Firm
ACR72585	NCBI	E.elig	<i>Eubacterium eligens</i>	Bact_Firm
AAQ46094	NCBI	E.faec	<i>Enterococcus faecium</i>	Bact_Firm
EEI61543	NCBI	E.faec	<i>Enterococcus faecium</i>	Bact_Firm
ZP_00602671	NCBI	E.faec	<i>Enterococcus faecium</i>	Bact_Firm
BAC14044	NCBI	O.ihey	<i>Oceanobacillus iheyensis</i>	Bact_Firm
ZP_02039674	NCBI	R.gnav	<i>Ruminococcus gnavus</i>	Bact_Firm
AAB08040	NCBI	T.therm	<i>Thermoanaerobacterium thermosulfurigenes</i>	Bact_Firm
ACB76612	NCBI	O.tera	<i>Opitutus terrae</i>	Bact_Chla
ACB77543	NCBI	O.tera	<i>Opitutus terrae</i>	Bact_Chla
ZP_01960177	NCBI	B.cac	<i>Bacteroides caccae</i>	Bact_Bac
EEF89217	NCBI	B.cel	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
EEF91172	NCBI	B.cel	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
EEF89217	NCBI	B.cell	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
EEF90911	NCBI	B.cell	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
ZP_02067915	NCBI	B.ova	<i>Bacteroides ovatus</i>	Bact_Bac

ZP_02068029	NCBI	B.ova	<i>Bacteroides ovatus</i>	Bact_Bac
ZP_02434149	NCBI	B.ster	<i>Bacteroides stercoris</i>	Bact_Bac
ZP_02437005	NCBI	B.ster	<i>Bacteroides stercoris</i>	Bact_Bac
AAO79228	NCBI	B.thet	<i>Bacteroides thetaiotomicron</i>	Bact_Bac
AAO79292	NCBI	B.thet	<i>Bacteroides thetaiotomicron</i>	Bact_Bac
ABR39657	NCBI	B.vulg	<i>Bacteroides vulgatus</i>	Bact_Bac
YP_001299279	NCBI	B.vulg	<i>Bacteroides vulgatus</i>	Bact_Bac
ACU62935	NCBI	C.pine	<i>Chitinophaga pinensis</i>	Bact_Bac
ABQ06132	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
ABQ07096	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
ABQ07123	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
ABQ07255	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
ABQ07264	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
ACU03376	NCBI	P.hepa	<i>Pedobacter heparinus</i>	Bact_Bac
ZP_02030629	NCBI	P.merd	<i>Parabacteroides merdae</i>	Bact_Bac
ACY49269	NCBI	R.mar	<i>Rhodothermus marinus</i>	Bact_Bac
EEO96050	NCBI	S.ling	<i>Spirosoma linguale</i>	Bact_Bac
EEO97598	NCBI	S.ling	<i>Spirosoma linguale</i>	Bact_Bac
EEO97813	NCBI	S.ling	<i>Spirosoma linguale</i>	Bact_Bac
ACO31589	NCBI	A.caps	<i>Acidobacterium capsulatum</i>	Bact_Acid
ABJ82244	NCBI	S.usit	<i>Solibacter usitatus</i>	Bact_Acid
ABJ88671	NCBI	S.usit	<i>Solibacter usitatus</i>	Bact_Acid
AAW84064	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
AAW84066	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ACL36472	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW01078	NCBI	C.maq	<i>Caldivirga maquilingensis</i>	Arch
CAJ19131	NCBI	Un.micr	<i>unidentified microorganism</i>	Unk

Dataset S02. List of sequences and species included in the PL3 Pectate Lyases analysis.

Accession numbers, source database, usual name (where applicable), abbreviated and full species name are indicated for every sequence retained in the phylogenetic analysis of PL3 Pectate Lyases. Sequences are sorted by taxonomic group then by species name. Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _CN for Cyst Nematode, _Aph for Aphelenchoidea; Bact: Bacteria, with the following suffixes: _Prot: Proteobacteria, _Firm: Firmicutes; _Act: Actinobacteria; Plant: Plant; Oomyc: Oomycete; Fungi: Fungi.

Accession	Source	Usual name	Species	Species long	Group
Mh_Contig329.2370..3481	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig1669.7002..6207	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig1669.8502..9658	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig331.11376..12399	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig841.11519..12588	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig841.8801..9875	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig400.8836..9978	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig711.8186..6817	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig418.19714..18783	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig418.1229..2522	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig338.76976..78669	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig338.80243..81506	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig338.101858..103127	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Minc18743	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc04661	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc05972	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc06289	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc12947	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc11928a	http://www.inra.fr/meloidogyne_incognita	Pel3	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAQ09004	NCBI	Pel1	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc03618	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01514	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc08824	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc09098	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc03619	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc08823	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01512	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc05810	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01516	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01810	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01513	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01522c	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01517	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc01515	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc14254	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc00379_Minc00378	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN

Minc11772	http://www.inra.fr/meloidogyne_incognita	Pel2	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc00380	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc14255	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc05811	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc13942	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc09783	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc16042	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAL66022	NCBI		M.jav	<i>Meloidogyne javanica</i>	PPN_RKN
ctg206645	http://www.sanger.ac.uk/sequencing/Globodera/pallida/		G.pal	<i>Globodera pallida</i>	PPN_CN
sca02986	http://www.sanger.ac.uk/sequencing/Globodera/pallida/		G.pal	<i>Globodera pallida</i>	PPN_CN
sca02674	http://www.sanger.ac.uk/sequencing/Globodera/pallida/		G.pal	<i>Globodera pallida</i>	PPN_CN
AAF80747	NCBI	Pel1	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37147	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37138	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37152	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37135	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37144	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37149	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
AAM21970	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37139	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37142	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37151	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37146	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37143	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37136	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
ACO37140	NCBI	Pel2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN
AAK08974	NCBI	Pel1	H.glyc	<i>Heterodera glycines</i>	PPN_CN
AAM74954	NCBI	Pel2	H.glyc	<i>Heterodera glycines</i>	PPN_CN
ABN14273	NCBI	Pel1	H.scha	<i>Heterodera schachtii</i>	PPN_CN
ABN14272	NCBI	Pel2	H.scha	<i>Heterodera schachtii</i>	PPN_CN
BAI44497	NCBI		A.ave	<i>Aphelenchus avenae</i>	PPN_Aph
BAI44499	NCBI		A.ave	<i>Aphelenchus avenae</i>	PPN_Aph
BAE48375	NCBI		B.mucr	<i>Bursaphelenchus mucronatus</i>	PPN_Aph
BAE48373	NCBI		B.mucr	<i>Bursaphelenchus mucronatus</i>	PPN_Aph
BAE48370	NCBI		B.xylo	<i>Bursaphelenchus xylophilus</i>	PPN_Aph
BAE48369	NCBI		B.xylo	<i>Bursaphelenchus xylophilus</i>	PPN_Aph
43398	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
80133	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
77024	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
71773	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
86169	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
87411	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
77029	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
74998	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc
46251	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	<i>Phytophthora ramorum</i>	Oomyc

39825	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
77022	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
44483	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
44445	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
44485	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
44487	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
82586	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html	P.ramo	<i>Phytophthora ramorum</i>	Oomyc
142501	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
144923	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
132156	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
132158	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
132101	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
140239	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
140241	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
109550	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
143281	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
132161	http://genome.jgi-psf.org/Physo1_1/Physo1_1.download.ftp.html	P.soj	<i>Phytophthora sojae</i>	Oomyc
EAW15258	NCBI	A.clav	<i>Aspergillus clavatus</i>	Fungi
EED50224	NCBI	A.flav	<i>Aspergillus flavus</i>	Fungi
EED45968	NCBI	A.flav	<i>Aspergillus flavus</i>	Fungi
EED57009	NCBI	A.flav	<i>Aspergillus flavus</i>	Fungi
EAL85355	NCBI	A.fum	<i>Aspergillus fumigatus</i>	Fungi
EAL86832	NCBI	A.fum	<i>Aspergillus fumigatus</i>	Fungi
EAL87824	NCBI	A.fum	<i>Aspergillus fumigatus</i>	Fungi
EAA64647	NCBI	A.nid	<i>Aspergillus nidulans</i>	Fungi
EAA67075	NCBI	A.nid	<i>Aspergillus nidulans</i>	Fungi
EAA63305	NCBI	A.nid	<i>Aspergillus nidulans</i>	Fungi
EAA58081	NCBI	A.nid	<i>Aspergillus nidulans</i>	Fungi
EAA58566	NCBI	A.nid	<i>Aspergillus nidulans</i>	Fungi
BAE66524	NCBI	A.ory	<i>Aspergillus oryzae</i>	Fungi
BAE55530	NCBI	A.ory	<i>Aspergillus oryzae</i>	Fungi
BAE64286	NCBI	A.ory	<i>Aspergillus oryzae</i>	Fungi
EAU30758	NCBI	A.terr	<i>Aspergillus terreus</i>	Fungi
EAU32829	NCBI	A.terr	<i>Aspergillus terreus</i>	Fungi
EAU32858	NCBI	A.terr	<i>Aspergillus terreus</i>	Fungi
EDN26806	NCBI	B.fuck	<i>Botryotinia fuckeliana</i>	Fungi
EDN30813	NCBI	B.fuck	<i>Botryotinia fuckeliana</i>	Fungi
EAU85565	NCBI	C.cine	<i>Coprinopsis cinerea</i>	Fungi
EAU85566	NCBI	C.cine	<i>Coprinopsis cinerea</i>	Fungi
EAQ92032	NCBI	C.glob	<i>Chaetomium globosum</i>	Fungi
EAQ91474	NCBI	C.glob	<i>Chaetomium globosum</i>	Fungi
EAQ84505	NCBI	C.glob	<i>Chaetomium globosum</i>	Fungi
EAQ90354	NCBI	C.glob	<i>Chaetomium globosum</i>	Fungi
XP_382562	NCBI	G.zea	<i>Gibberella zeae</i>	Fungi
XP_390180	NCBI	G.zea	<i>Gibberella zeae</i>	Fungi

AAV66343	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
XP_385040	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
ABT35769	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
XP_383153	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
XP_383889	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
XP_384085	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
XP_391270	NCBI	G.zea	<i>Gibberella zaeae</i>	Fungi
EDJ96534	NCBI	M.gris	<i>Magnaporthe grisea</i>	Fungi
XP_369589	NCBI	M.gris	<i>Magnaporthe grisea</i>	Fungi
XP_328882	NCBI	N.crass	<i>Neurospora crassa</i>	Fungi
EEU34673	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
AAA33338	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
AAA87382	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU44672	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU45012	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
AAA87383	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU34414	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU38652	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU43672	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
AAC49420	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU36069	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU35915	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
EEU38641	NCBI	N.haem	<i>Nectria haematococca</i>	Fungi
CAP97876	NCBI	P.chr	<i>Penicillium chrysogenum</i>	Fungi
EAT84565	NCBI	P.nodo	<i>Phaeosphaeria nodorum</i>	Fungi
EEY23693	NCBI	V.alb	<i>Verticillium albo-atrum</i>	Fungi
EEY23761	NCBI	V.alb	<i>Verticillium albo-atrum</i>	Fungi
EEY23679	NCBI	V.alb	<i>Verticillium albo-atrum</i>	Fungi
EEY18959	NCBI	V.alb	<i>Verticillium albo-atrum</i>	Fungi
EEY23009	NCBI	V.dah	<i>Verticillium dahliae</i>	Fungi
EEY15367	NCBI	V.dah	<i>Verticillium dahliae</i>	Fungi
XP_382562	NCBI	F.oxy	<i>Fusarium oxysporum</i>	Fungi
AAC64368	NCBI	F.oxy	<i>Fusarium oxysporum</i>	Fungi
EAW20237	NCBI	N.fisch	<i>Neosartorya fischeri</i>	Fungi
EAW19695	NCBI	N.fisch	<i>Neosartorya fischeri</i>	Fungi
EAW23636	NCBI	N.fisch	<i>Neosartorya fischeri</i>	Fungi
CAP65892	NCBI	P.anse	<i>Podospora anserina</i>	Fungi
CAP61540	NCBI	P.anse	<i>Podospora anserina</i>	Fungi
EDU43763	NCBI	P.trit	<i>Pyrenophora tritici-repentis</i>	Fungi
EDU47604	NCBI	P.trit	<i>Pyrenophora tritici-repentis</i>	Fungi
EDU40146	NCBI	P.trit	<i>Pyrenophora tritici-repentis</i>	Fungi
ACE85330	NCBI	C.jap	<i>Cellvibrio japonicus</i>	Bact_Prot
ACE83519	NCBI	C.jap	<i>Cellvibrio japonicus</i>	Bact_Prot
CAN93573	NCBI	S.cell	<i>Sorangium cellulosum</i>	Bact_Prot
ABD79872	NCBI	S.deg	<i>Saccharophagus degradans</i>	Bact_Prot

CAB40884	NCBI	B.sp.	<i>Bacillus sp</i>	Bact_Firm
ACU36508	NCBI	A.mir	<i>Actinosynnema mirum</i>	Bact_Act
ACU36500	NCBI	A.mir	<i>Actinosynnema mirum</i>	Bact_Act
ACU38032	NCBI	A.mir	<i>Actinosynnema mirum</i>	Bact_Act
EEN37459	NCBI	C.flav	<i>Cellulomonas flavigena</i>	Bact_Act
CAQ02326	NCBI	C.mich	<i>Clavibacter michiganensis</i>	Bact_Act
CAN00065	NCBI	C.mich	<i>Clavibacter michiganensis</i>	Bact_Act
CAN00055	NCBI	C.mich	<i>Clavibacter michiganensis</i>	Bact_Act
ABW11736	NCBI	Fra.sp	<i>Frankia sp.</i>	Bact_Act
ACV08000	NCBI	J.dent	<i>Jonesia denitrificans</i>	Bact_Act
CAJ90085	NCBI	S.ambo	<i>Streptomyces ambofaciens</i>	Bact_Act
BAC74094	NCBI	S.aver	<i>Streptomyces avermitilis</i>	Bact_Act
CAC13062	NCBI	S.coel	<i>Streptomyces coelicolor</i>	Bact_Act
CAM03631	NCBI	S.eryt	<i>Saccharopolyspora erythraea</i>	Bact_Act
ABA02177	NCBI	un.bac	<i>Uncultured bacterium</i>	Bact

Dataset S03. List of sequences and species included in the GH43 candidate arabinanase analysis.

Accession numbers, source database, usual name (where applicable), abbreviated and full species names are indicated for every sequence retained in the phylogenetic analysis of GH43 candidate arabinanases. Sequences are sorted by taxonomic group then by species name. Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _CN for Cyst Nematode; Bact: Bacteria, with the following suffixes: _Prot: Proteobacteria, _Firm: Firmicutes, _Dict: Dictyoglomi, _Chla: Chlamydiae / Verrucomicrobia, _Bac: Bacteroidetes, _Act: Actinobacteria,

Accession	Source	Usual name	Species	Species long	Group
Mh_Contig1558.5336..7500	http://www.hapla.org		M.hap	Meloidogyne hapla	PPN_RKN
Minc10639	http://www.inra.fr/meloidogyne_incognita		M.inc	Meloidogyne incognita	PPN_RKN
Minc05403	http://www.inra.fr/meloidogyne_incognita		M.inc	Meloidogyne incognita	PPN_RKN
sca01805	http://www.sanger.ac.uk/sequencing/Globodera/pallida/		G.pal	Globodera pallida	PPN_CN
75229	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	Phytophthora ramorum	Oomyc
75228	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	Phytophthora ramorum	Oomyc
75238	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	Phytophthora ramorum	Oomyc
71648	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.download.ftp.html		P.ramo	Phytophthora ramorum	Oomyc
109619	http://genome.jgi-psf.org/Phys01_1/Phys01_1.download.ftp.html		P.soj	Phytophthora sojae	Oomyc
144181	http://genome.jgi-psf.org/Phys01_1/Phys01_1.download.ftp.html		P.soj	Phytophthora sojae	Oomyc
144175	http://genome.jgi-psf.org/Phys01_1/Phys01_1.download.ftp.html		P.soj	Phytophthora sojae	Oomyc
144182	http://genome.jgi-psf.org/Phys01_1/Phys01_1.download.ftp.html		P.soj	Phytophthora sojae	Oomyc
AAG27441	NCBI		A.acu	Aspergillus aculeatus	Fungi
EAW08955	NCBI		A.clav	Aspergillus clavatus	Fungi
EAW09988	NCBI		A.clav	Aspergillus clavatus	Fungi
EED51687	NCBI		A.flav	Aspergillus flavus	Fungi
EED48144	NCBI		A.flav	Aspergillus flavus	Fungi
EAL84189	NCBI		A.fum	Aspergillus fumigatus	Fungi
EAL92126	NCBI		A.fum	Aspergillus fumigatus	Fungi
EAA64639	NCBI		A.nid	Aspergillus nidulans	Fungi
CAK44404	NCBI		A.nig	Aspergillus niger	Fungi
AAA32682	NCBI		A.nig	Aspergillus niger	Fungi
CAK49041	NCBI		A.nig	Aspergillus niger	Fungi
AAA32682	NCBI		A.nig	Aspergillus niger	Fungi
BAE64437	NCBI		A.oryz	Aspergillus oryzae	Fungi
BAE62189	NCBI		A.oryz	Aspergillus oryzae	Fungi
EDN31415	NCBI		B.fuck	Botryotinia fuckeliana	Fungi
CAJ81227	NCBI		B.rhod	Botryosphaeria rhodina	Fungi
EAQ83141	NCBI		C.glob	Chaetomium globosum	Fungi
EAQ90215	NCBI		C.glob	Chaetomium globosum	Fungi
XP_383178	NCBI		G.zeae	Gibberella zeae	Fungi
EAW21105	NCBI		N.fisc	Neosartorya fischeri	Fungi
EEU39838	NCBI		N.haem	Nectria haematococca	Fungi
CAP79831	NCBI		P.chry	Phanerochaete chrysosporium	Fungi
BAD15018	NCBI		P.chry	Phanerochaete chrysosporium	Fungi
CAP92661	NCBI		P.chry	Phanerochaete chrysosporium	Fungi
BAD89094	NCBI		P.chry	Phanerochaete chrysosporium	Fungi

CAP86902	NCBI	P.chry	Phanerochaete chrysosporium	Fungi
EDU42696	NCBI	P.trit	Pyrenophora tritici-repentis	Fungi
EDN97397	NCBI	S.scle	Sclerotinia sclerotiorum	Fungi
EEY16682	NCBI	V.albo	Verticillium albo-atrum	Fungi
ACE84667	NCBI	C.jap	Cellvibrio japonicus	Bact_Prot
ACK46658	NCBI	S.balt	Shewanella baltica	Bact_Prot
CAN91643	NCBI	S.cell	Sorangium cellulosum	Bact_Prot
ABD80276	NCBI	S.degr	Saccharophagus degradans	Bact_Prot
ABD80048	NCBI	S.degr	Saccharophagus degradans	Bact_Prot
ABP75772	NCBI	S.putr	Shewanella putrefaciens	Bact_Prot
ABM24796	NCBI	S.sp.	Shewanella sp.	Bact_Prot
ABI42969	NCBI	S.sp.	Shewanella sp.	Bact_Prot
ABK48312	NCBI	S.sp.	Shewanella sp.	Bact_Prot
ABI39068	NCBI	S.sp.	Shewanella sp.	Bact_Prot
ABS74945	NCBI	B.amyl	Bacillus amyloliquefaciens	Bact_Firm
AAU40201	NCBI	B.lich	Bacillus licheniformis	Bact_Firm
AAU41895	NCBI	B.lich	Bacillus licheniformis	Bact_Firm
BAA20372	NCBI	B.subt	Bacillus subtilis	Bact_Firm
CAA99586	NCBI	B.subt	Bacillus subtilis	Bact_Firm
ABP67152	NCBI	C.sacc	Caldicellulosiruptor saccharolyticus	Bact_Firm
ACE73676	NCBI	G.stea	Geobacillus stearothermophilus	Bact_Firm
BAB64339	NCBI	G.therm	Geobacillus thermodenitrificans	Bact_Firm
ACK43001	NCBI	D.turg	Dictyoglomus turgidum	Bact_Dict
ACB74675	NCBI	O.terr	Opitutus terrae	Bact_Chla
EEF90586	NCBI	B.cell	Bacteroides cellulosilyticus	Bact_Bac
EEF90587	NCBI	B.cell	Bacteroides cellulosilyticus	Bact_Bac
EEF90382	NCBI	B.cell	Bacteroides cellulosilyticus	Bact_Bac
YP_210029	NCBI	B.frag	Bacteroides fragilis	Bact_Bac
CAH06067	NCBI	B.frag	Bacteroides fragilis	Bact_Bac
BAD47093	NCBI	B.frag	Bacteroides fragilis	Bact_Bac
ZP_02065207	NCBI	B.ovat	Bacteroides ovatus	Bact_Bac
AAO78006	NCBI	B.thet	Bacteroides thetaiotaomicron	Bact_Bac
AAO78622	NCBI	B.thet	Bacteroides thetaiotaomicron	Bact_Bac
EDO55411	NCBI	B.unif	Bacteroides uniformis	Bact_Bac
YP_001299513	NCBI	B.vulg	Bacteroides vulgatus	Bact_Bac
ABR39891	NCBI	B.vulg	Bacteroides vulgatus	Bact_Bac
ACT95486	NCBI	D.ferm	Dyadobacter fermentans	Bact_Bac
CAL65671	NCBI	G.fors	Gramella forsetii	Bact_Bac
EDN84672	NCBI	P.merd	Parabacteroides merdae	Bact_Bac
ACU37633	NCBI	A.miru	Actinosynnema mirum	Bact_Act
EEN37136	NCBI	C.flav	Cellulomonas flavigena	Bact_Act
ABS01569	NCBI	K.radi	Kineococcus radiotolerans	Bact_Act
EEK32730	NCBI	N.dass	Nocardiopsis dassonvillei	Bact_Act
CAB92901	NCBI	S.coel	Streptomyces coelicolor	Bact_Act
ACZ20417	NCBI	S.kedd	Sanguibacter keddieii	Bact_Act
ACY98358	NCBI	T.curv	Thermomonospora curvata	Bact_Act

ABJ85402	NCBI	S.usit	Solibacter usitatus	Bact_Acid
ABJ83973	NCBI	S.usit	Solibacter usitatus	Bact_Acid
ACM91039	NCBI	Un.Bac	Uncultured bacterium	Bact

Dataset S04. List of sequences and species included in the GH5 cellulases phylogenetic analysis.

Accession numbers, source database, usual name (where applicable), abbreviated and full species name are indicated for every sequence retained in the phylogenetic analysis of GH5 cellulases. Sequences are sorted by taxonomic group then by species name. Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _CN for Cyst Nematode, _Aph for Aphelenchidae, _Rad for Radopholinae, _Prat for Pratylenchidae, _Ang for Anguinidae; Bact: Bacteria, with the following suffixes: _Prot: Proteobacteria, _Firm: Firmicutes, _Fibr: Fibrobacteres, _Bac: Bacteroidetes, _Act: Actinobacteria; Fungi: Fungi; Ins: Insect; Unk: unidentified organism. Nematode cellulases that possess a CBM2 module are indicated in the last column. Note that sequence with accession number AAC49731 and annotated as being from the fungus *Orpinomyces joyonii* is certainly a bacterial contaminant as it shares 95% identity with GH5 cellulases from *Fibrobacter succinogenes* and does not significantly match any other fungal sequence. It was thus considered as a bacterial

Accession	Source	Usual name	Species	Species long	Group	CBM2
ABH12705	NCBI		Unk	<i>Unknown</i>	Unk	
AAM56798	NCBI		Unk	<i>Unknown</i>	Unk	
CAJ19149	NCBI		Unk	<i>unidentified microorganism</i>	Unk	
BAD90558	NCBI		S.leid	<i>Spirotrichonympha leidyi</i>	Prot	
BAF57458	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57334	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57460	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57330	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57290	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57361	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57292	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57332	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57328	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57329	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57327	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
BAF57341	NCBI		Symb.prot	<i>uncultured symbiotic protist of Termite</i>	Prot	
Mh_Contig802.24595..26731	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig344.7550..8690	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig344.4481..6245	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	CBM2
Mh_Contig392.21227..22975	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	CBM2
Minc09446a	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
AAR37375	NCBI	ENG4	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
AAK21882	NCBI	ENG1a	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc10405	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc18711	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
AAK21883	NCBI	ENG2	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc19090a	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	

AAR37374	NCBI		ENG3	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc00168	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc03286	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc14048	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc18310_Minc18311	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc13614	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
AAK21881	NCBI		ENG2	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc02089a	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc00169a	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc03287	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc14047a	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc13221a	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
AAD45868	NCBI		ENG1	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc09298a	http://www.inra.fr/meloidogyne_incognita			M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
CAJ77137	NCBI		ENG3	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	CBM2
ABV54448	NCBI		ENG2	R.sim	<i>Radopholus similis</i>	PPN_Rad	
ABV54449	NCBI		ENG3	R.sim	<i>Radopholus similis</i>	PPN_Rad	
ABV54446	NCBI		ENG1a	R.sim	<i>Radopholus similis</i>	PPN_Rad	CBM2
ABV54447	NCBI		ENG1b	R.sim	<i>Radopholus similis</i>	PPN_Rad	CBM2
ABX79356	NCBI		ENG1	P.coff	<i>Pratylenchus coffeae</i>	PPN_Prat	CBM2
BAB68523	NCBI		ENG2	P.pen	<i>Pratylenchus penetrans</i>	PPN_Prat	CBM2
BAB68522	NCBI		ENG1	P.pen	<i>Pratylenchus penetrans</i>	PPN_Prat	CBM2
AAN03645	NCBI		ENG3	G.rost	<i>Globodera rostochiensis</i>	PPN_CN	
AAN03647	NCBI		ENG4	G.rost	<i>Globodera rostochiensis</i>	PPN_CN	
AAC63988	NCBI		ENG1	G.rost	<i>Globodera rostochiensis</i>	PPN_CN	CBM2
AAC48341	NCBI		ENG2	G.rost	<i>Globodera rostochiensis</i>	PPN_CN	
AAD56392	NCBI		ENG1	G.tabata	<i>Globodera tabacum solanacearum</i>	PPN_CN	CBM2
AAN32884	NCBI		ENG5	H.glyc	<i>Heterodera glycines</i>	PPN_CN	
AAC15707	NCBI		ENG1	H.glyc	<i>Heterodera glycines</i>	PPN_CN	CBM2
AAC48326	NCBI		ENG2	H.glyc	<i>Heterodera glycines</i>	PPN_CN	
AAK85303	NCBI		ENG4	H.glyc	<i>Heterodera glycines</i>	PPN_CN	
AAC33848	NCBI		ENG3	H.glyc	<i>Heterodera glycines</i>	PPN_CN	
AAM50039	NCBI		ENG6	H.glyc	<i>Heterodera glycines</i>	PPN_CN	
CAC12959	NCBI		ENG2	H.scha	<i>Heterodera schachtii</i>	PPN_CN	
BAI44495	NCBI		ENG1	A.ave	<i>Aphelenchus avenae</i>	PPN_Aph	
BAI44493	NCBI		ENG2	A.ave	<i>Aphelenchus avenae</i>	PPN_Aph	
ABY52965	NCBI		ENG1	D.afr	<i>Ditylenchus africanus</i>	PPN_Ang	CBM2
AAX18655	NCBI			A.germ	<i>Apriona germari</i>	Ins	

BAB86867	NCBI	P.hila	<i>Psacothea hilaris</i>	Ins
AAC49731	NCBI	O.joy	<i>Orpinomyces joyonii</i>	Fungi*
ACE84076	NCBI	C.japo	<i>Cellvibrio japonicus</i>	Bact_Prot
AAB61462	NCBI	C.mixt	<i>Cellvibrio mixtus</i>	Bact_Prot
AAF18152	NCBI	E.chr	<i>Erwinia chrysanthemi</i>	Bact_Prot
P07103	NCBI	E.chr	<i>Erwinia chrysanthemi</i>	Bact_Prot
ABC30636	NCBI	H.chej	<i>Hahella chejuensis</i>	Bact_Prot
ABG39929	NCBI	P.atla	<i>Pseudoalteromonas atlantica</i>	Bact_Prot
CAG74882	NCBI	P.atro	<i>Pectobacterium atrosepticum</i>	Bact_Prot
AAC02964	NCBI	P.car	<i>Pectobacterium carotovorum</i>	Bact_Prot
CAA53592	NCBI	P.car	<i>Pectobacterium carotovorum</i>	Bact_Prot
CAA55823	NCBI	P.car	<i>Pectobacterium carotovorum</i>	Bact_Prot
AAC37033	NCBI	P.car	<i>Pectobacterium carotovorum</i>	Bact_Prot
CAA76775	NCBI	P.hal	<i>Pseudoalteromonas haloplanktis</i>	Bact_Prot
AAS58467	NCBI	Ps.sp	<i>Pseudomonas sp.</i>	Bact_Prot
BAB79288	NCBI	Ps.sp	<i>Pseudomonas sp.</i>	Bact_Prot
ACN62172	NCBI	Ps.sp	<i>Pseudomonas sp.</i>	Bact_Prot
CAN91626	NCBI	S.cell	<i>Sorangium cellulosum</i>	Bact_Prot
CAN95073	NCBI	S.cell	<i>Sorangium cellulosum</i>	Bact_Prot
ABD82260	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot
ABD82260	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot
ABD82496	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot
ABD81754	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot
ABD82494	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot
ABS72374	NCBI	T.turn	<i>Teredinibacter turnerae</i>	Bact_Prot
CAB01405	NCBI	A.ther	<i>Anaerocellum thermophilum DSM 6725</i>	Bact_Firm
AAC19169	NCBI	B.agar	<i>Bacillus agaradhaerens</i>	Bact_Firm
CAE11243	NCBI	B.amyl	<i>Bacillus amyloliquefaciens</i>	Bact_Firm
AAL99668	NCBI	B.amyl	<i>Bacillus amyloliquefaciens</i>	Bact_Firm
AAA20893	NCBI	B.fibr	<i>Butyrivibrio fibrisolvens</i>	Bact_Firm
CAJ70714	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
AAU40977	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
AAP51020	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
CAJ70710	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
CAE82178	NCBI	B.lich	<i>Bacillus licheniformis</i>	Bact_Firm
BAA31712	NCBI	B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAV34758	NCBI	B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAA22304	NCBI	B.sp.	<i>Bacillus sp.</i>	Bact_Firm

BAB19360	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAA22299	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAA22301	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
BAB62295	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAU08303	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAX54913	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
CAA83942	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAC02536	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
CAB59165	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAC43478	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAK73277	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAT97263	NCBI		B.sp.	<i>Bacillus sp.</i>	Bact_Firm
AAA22496	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ABS70711	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAN07019	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAA22307	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ABS70712	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAO63626	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
CAA47429	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAK39540	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
BAA00859	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ACK38261	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ACI15227	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
CAA82317	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
CAA28392	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ACM68436	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ABG78039	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAK94871	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAZ22322	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
ABV45393	NCBI		B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAK78683	NCBI		C.acet	<i>Clostridium acetobutylicum</i>	Bact_Firm
ACL74811	NCBI		C.cell	<i>Clostridium cellulolyticum</i>	Bact_Firm
ZP_02206208	NCBI		C.eut	<i>Coprococcus eutactus</i>	Bact_Firm
BAA12826	NCBI		C.jos	<i>Clostridium josui</i>	Bact_Firm
AAA23230	NCBI		C.sacc	<i>Clostridium saccharobutylicum</i>	Bact_Firm
ABP66297	NCBI		C.sacc	<i>Clostridium saccharobutylicum</i>	Bact_Firm
ZP_02075188	NCBI		C.sp.	<i>Clostridium sp.</i>	Bact_Firm
AAB40891	NCBI		Cl.cel	<i>Clostridium cellulovorans</i>	Bact_Firm

ACR73731	NCBI	E.elig	<i>Eubacterium eligens</i>	Bact_Firm
ACR74405	NCBI	E.rec	<i>Eubacterium rectale</i>	Bact_Firm
ZP_02421845	NCBI	E.sir	<i>Eubacterium siraeum</i>	Bact_Firm
CAA01935	NCBI	P.laut	<i>Paenibacillus laetus</i>	Bact_Firm
ABV08875	NCBI	P.poly	<i>Paenibacillus polymyxa</i>	Bact_Firm
BAF62085	NCBI	P.sp.	<i>Paenibacillus sp.</i>	Bact_Firm
BAF62087	NCBI	P.sp.	<i>Paenibacillus sp.</i>	Bact_Firm
BAF62084	NCBI	P.sp.	<i>Paenibacillus sp.</i>	Bact_Firm
BAF62086	NCBI	P.sp.	<i>Paenibacillus sp.</i>	Bact_Firm
BAA32286	NCBI	R.alb	<i>Ruminococcus albus</i>	Bact_Firm
AAM23649	NCBI	T.teng	<i>Thermoanaerobacter tengcongensis</i>	Bact_Firm
AAC06197	NCBI	F.suc	<i>Fibrobacter succinogenes</i>	Bact_Fibr
ABU45500	NCBI	F.suc	<i>Fibrobacter succinogenes</i>	Bact_Fibr
EEF89984	NCBI	B.cell	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
ZP_02065640	NCBI	B.ovat	<i>Bacteroides ovatus</i>	Bact_Bac
EDO55904	NCBI	B.uni	<i>Bacteroides uniformis</i>	Bact_Bac
ABG59366	NCBI	C.hutc	<i>Cytophaga hutchinsonii</i>	Bact_Bac
ABG58383	NCBI	C.hutc	<i>Cytophaga hutchinsonii</i>	Bact_Bac
ABQ07945	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac
AAC06196	NCBI	A.sp.	<i>Actinomyces sp. 40</i>	Bact_Act
CAC16970	NCBI	S.coel	<i>Streptomyces coelicolor</i>	Bact_Act
CAM03275	NCBI	S.eryt	<i>Saccharopolyspora erythraea</i>	Bact_Act
BAG18170	NCBI	S.gris	<i>Streptomyces griseus</i>	Bact_Act
BAC65342	NCBI	S.gris	<i>Streptomyces griseus</i>	Bact_Act
P27035	NCBI	S.liv	<i>Streptomyces lividans</i>	Bact_Act
EEP04998	NCBI	S.nass	<i>Stackebrandtia nassauensis</i>	Bact_Act
ABP56852	NCBI	S.trop	<i>Salinisporea tropica</i>	Bact_Act
ABP54735	NCBI	S.trop	<i>Salinisporea tropica</i>	Bact_Act
AAZ54939	NCBI	T.fus	<i>Thermobifida fusca YX</i>	Bact_Act
AAP04424	NCBI	MaBac	<i>psychrophilic marine bacterium</i>	Bact
ABI94085	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39357	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39342	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39355	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39337	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39356	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39336	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39358	NCBI	UnBac	<i>uncultured bacterium</i>	Bact

ABA42184	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABW39350	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABI94086	NCBI	UnBac	<i>uncultured bacterium</i>	Bact
ABA42185	NCBI	UnBac	<i>uncultured bacterium</i>	Bact

* possible bacterial contamination (95% identity with *Fibrobacter succinogenes* GH5 cellulases)

Dataset S05. List of sequences and species included in the GH5 xylanases phylogenetic analysis.

Accession numbers, source database, usual name (when applicable), abbreviated and full species name are indicated for every sequence retained in the phylogenetic analysis of GH5 xylanases. Sequences are sorted by taxonomic group then by species name. Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _Rad for Radopholinae; Bact: Bacteria, with the following suffixes: _Ther: Thermotogae, _Tene: Tenericutes, _Prot: Proteobacteria, _Firm: Firmicutes, _Bac: Bacteroidetes, _Acid: Acidobacteria; Fungi: Fungi; Met: Metazoa.

Accession	Source	Usual name	Species	Species long	Group	GH30
MA00274	http://www.nematode.net/		M.are	<i>Meloidogyne arenaria</i>	PPN_RKN	
MC03416	http://www.nematode.net/		M.chit	<i>Meloidogyne chitwoodi</i>	PPN_RKN	
MC02136	http://www.nematode.net/		M.chit	<i>Meloidogyne chitwoodi</i>	PPN_RKN	
Mh_Contig2188.1417..3259	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Minc18739	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc18650	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc11164	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc18912	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc09178	http://www.inra.fr/meloidogyne_incognita	Xyl3	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc03215	http://www.inra.fr/meloidogyne_incognita	Xyl2	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
AAF37276	NCBI	Xyl1	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
MJ04776	http://www.nematode.net/		M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
MJ04695	http://www.nematode.net/		M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
ABZ78968	NCBI	Xyl1	R.sim	<i>Radopholus similis</i>	PPN_Rad	
XP_001897915	NCBI		B.mala	<i>Brugia malayi</i>	Met	*
CAC35807	NCBI		C.ele	<i>Caenorhabditis elegans</i>	Met	*
AAB66060	NCBI		C.ele	<i>Caenorhabditis elegans</i>	Met	*
NP_494207	NCBI		C.ele	<i>Caenorhabditis elegans</i>	Met	*
XP_001275546	NCBI		A.clav	<i>Aspergillus clavatus</i>	Fungi	
XP_001225956	NCBI		C.glob	<i>Chaetomium globosum</i>	Fungi	
AAP64786	NCBI		H.jec	<i>Hypocrea jecorina</i>	Fungi	
XP_001904921	NCBI		P.ans	<i>Podospora anserina</i>	Fungi	
CAP64828	NCBI		P.ans	<i>Podospora anserina</i>	Fungi	
EAT85735	NCBI		P.nod	<i>Phaeosphaeria nodorum</i>	Fungi	
EEY17075	NCBI		V.albo	<i>Verticillium albo-atrum</i>	Fungi	
ABS60162	NCBI		F.nodo	<i>Fervidobacterium nodosum</i>	Bact_Ther	*
ABX81342	NCBI		A.laid	<i>Acholeplasma laidlawii</i>	Bact_Tene	*
BAA13641	NCBI		A.punc	<i>Aeromonas punctata</i>	Bact_Prot	
AAB63573	NCBI		A.punc	<i>Aeromonas punctata</i>	Bact_Prot	

ACE83181	NCBI	C.japo	<i>Cellvibrio japonicus</i>	Bact_Prot
ACE85688	NCBI	C.japo	<i>Cellvibrio japonicus</i>	Bact_Prot *
AAB53151	NCBI	E.chrys	<i>Erwinia chrysanthemi</i>	Bact_Prot
AAL16415	NCBI	E.chrys	<i>Erwinia chrysanthemi</i>	Bact_Prot
ACY18721	NCBI	H.och	<i>Haliangium ochraceum</i>	Bact_Prot *
ABD68847	NCBI	R.ferr	<i>Rhodoferax ferrireducens</i>	Bact_Prot *
ABD82251	NCBI	S.degr	<i>Saccharophagus degradans</i>	Bact_Prot *
ACR13841	NCBI	T.turn	<i>Teredinibacter turnerae</i>	Bact_Prot
ACR14618	NCBI	T.turn	<i>Teredinibacter turnerae</i>	Bact_Prot
AAM35821	NCBI	X.axo	<i>Xanthomonas axonopodis</i>	Bact_Prot
AAM40172	NCBI	X.camp	<i>Xanthomonas campestris</i>	Bact_Prot
AAY50417	NCBI	X.camp	<i>Xanthomonas campestris</i>	Bact_Prot
CAP52858	NCBI	X.camp	<i>Xanthomonas campestris</i>	Bact_Prot
CAJ22596	NCBI	X.camp	<i>Xanthomonas campestris</i>	Bact_Prot
CAP52634	NCBI	X.camp	<i>Xanthomonas campestris</i>	Bact_Prot *
BAE70175	NCBI	X.oryz	<i>Xanthomonas oryzae</i>	Bact_Prot
ACD57852	NCBI	X.oryz	<i>Xanthomonas oryzae</i>	Bact_Prot
AAW76872	NCBI	X.oryz	<i>Xanthomonas oryzae</i>	Bact_Prot
CAE11246	NCBI	B.amyl	<i>Bacillus amyloliquefaciens</i>	Bact_Firm
ABS74297	NCBI	B.amyl	<i>Bacillus amyloliquefaciens</i>	Bact_Firm *
ABV62419	NCBI	B.pum	<i>Bacillus pumilus</i>	Bact_Firm
CAA97612	NCBI	B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAN07016	NCBI	B.subt	<i>Bacillus subtilis</i>	Bact_Firm
AAK76864	NCBI	C.acet	<i>Clostridium acetobutylicum</i>	Bact_Firm
AAK76863	NCBI	C.acet	<i>Clostridium acetobutylicum</i>	Bact_Firm
ACA57135	NCBI	C.botu	<i>Clostridium botulinum</i>	Bact_Firm *
ABS41885	NCBI	C.botu	<i>Clostridium botulinum</i>	Bact_Firm *
ACA44855	NCBI	C.botu	<i>Clostridium botulinum</i>	Bact_Firm *
ACL75598	NCBI	C.cell	<i>Clostridium cellulolyticum</i>	Bact_Firm
ACL75030	NCBI	C.cell	<i>Clostridium cellulolyticum</i>	Bact_Firm
ABN54208	NCBI	C.ther	<i>Clostridium thermocellum</i>	Bact_Firm
BAB39494	NCBI	R.albu	<i>Ruminococcus albus</i>	Bact_Firm
AAM25490	NCBI	T.teng	<i>Thermoanaerobacter tengcongensis</i>	Bact_Firm *
CAA11771	NCBI	A.orie	<i>Amycolatopsis orientalis</i>	Bact_Bac
EDM21323	NCBI	B.cac	<i>Bacteroides caccae</i>	Bact_Bac
ZP_03681316	NCBI	B.cell	<i>Bacteroides cellulosilyticus</i>	Bact_Bac
ZP_02066435	NCBI	B.ovat	<i>Bacteroides ovatus</i>	Bact_Bac

AAO78132	NCBI	B.thet	<i>Bacteroides thetaiotaomicron</i>	Bact_Bac
ABR38295	NCBI	B.vulg	<i>Bacteroides vulgatus</i>	Bact_Bac
YP_003113546	NCBI	C.acid	<i>Catenulispora acidiphila</i>	Bact_Bac
ACU71659	NCBI	C.acid	<i>Catenulispora acidiphila</i>	Bact_Bac
ACU73817	NCBI	C.acid	<i>Catenulispora acidiphila</i>	Bact_Bac
ABG59305	NCBI	C.hutch	<i>Cytophaga hutchinsonii</i>	Bact_Bac
ACU64137	NCBI	C.pine	<i>Chitinophaga pinensis</i>	Bact_Bac *
ACU61803	NCBI	C.pine	<i>Chitinophaga pinensis</i>	Bact_Bac *
ACT96791	NCBI	D.ferm	<i>Dyadobacter fermentans</i>	Bact_Bac *
ABQ04594	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac *
ABQ04937	NCBI	F.john	<i>Flavobacterium johnsoniae</i>	Bact_Bac *
ZP_04488653	NCBI	S.ling	<i>Spirosoma linguale</i>	Bact_Bac *
ACO32517	NCBI	A.caps	<i>Acidobacterium capsulatum</i>	Bact_Acid
ACO32413	NCBI	A.caps	<i>Acidobacterium capsulatum</i>	Bact_Acid *
ABF43153	NCBI	C.kori	<i>Candidatus Koribacter versatilis</i>	Bact_Acid *

* these proteins belong to the GH30 family which shares significant sequence similarity with GH5 xylanases but that is structurally and functionally unrelated
(i.e. no xylanase activity reported in this family)

Dataset S06. List of sequences and species included in the EXPN expansin-like analysis.

Accession numbers, source database, usual name (when applicable), abbreviated and full species name are indicated for every sequence retained in the phylogenetic analysis of EXPN expansin-like proteins. Sequences are sorted by species abbreviated name. Sequences are sorted by taxonomic group then by species name.

Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _CN for cyst nematode, _Ang for Anguinidae, _Dory for Dorylaimida, _Aph for Aphelenchoidea; Bact: Bacteria, with the following suffix: _Act: Actinobacteria; Fungi: Fungi; Amoe: Amoeba. Sequences

Accession	Source	Usual name	Species	Species long	Group	CBM2
Mh_Contig554.11824..12462	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig1015.18978..20118	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	CBM2
Mh_Contig790.11104..12567	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	CBM2
Mh_Contig1905.1608..2586	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig2646.6246..7094	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig1277.27768..28838	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Mh_Contig1905.330..998	http://www.hapla.org		M.hap	<i>Meloidogyne hapla</i>	PPN_RKN	
Minc10365	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc10366	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc00158	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc10987	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc10988	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc12725	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc14599	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc03214	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc07955	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc07960	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc11468	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc14116	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	CBM2
Minc14117	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc18856	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
Minc04469	http://www.inra.fr/meloidogyne_incognita		M.inc	<i>Meloidogyne incognita</i>	PPN_RKN	
CAP59535	NCBI	Ap1	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
CAP59536	NCBI	Ap2	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
CAP59537	NCBI	Map1	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
CAP59538	NCBI	Map2	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN	
CV509243	NCBI		X.ind	<i>Xiphinema index</i>	PPN_Dory	
ACN58322	NCBI	expansin B2 (ExpB2)	G.ros	<i>Globodera rostochiens</i>	PPN_CN	
CAC84564	NCBI	Expb2	G.ros	<i>Globodera rostochiens</i>	PPN_CN	
CAC83611	NCBI	expansin (Expb1)	G.ros	<i>Globodera rostochiens</i>	PPN_CN	CBM2

BAG16534	NCBI		B.mucr	<i>Bursaphelenchus mucronatus</i>	PPN_Aph
ABV60414	NCBI		B.xyl	<i>Bursaphelenchus xylophilus</i>	PPN_Aph
BAG16533	NCBI		B.xyl	<i>Bursaphelenchus xylophilus</i>	PPN_Aph
CL496 (FE924410 + FE922616)	NCBI		D.afr	<i>Ditylenchus africanus</i>	PPN_Ang
EAL91558	NCBI		A.fum	<i>Aspergillus fumigatus Af293</i>	Fungi
BAE66341	NCBI		A.ory	<i>Aspergillus oryzae RIB40</i>	Fungi
EDN21755	NCBI		B.fuc	<i>Botryotinia fuckeliana T4</i>	Fungi
EAL04023	NCBI		C.albi	<i>Candida albicans</i>	Fungi
EAU93049	NCBI		C.cin	<i>Coprinopsis cinerea Okayama 7</i>	Fungi
EAU88038	NCBI		C.cin	<i>Coprinopsis cinerea Okayama 7</i>	Fungi
EAU88039	NCBI		C.cin	<i>Coprinopsis cinerea Okayama 7</i>	Fungi
EAU90886	NCBI		C.cin	<i>Coprinopsis cinerea Okayama 7</i>	Fungi
EDR15021	NCBI		L.bico	<i>Laccaria bicolor</i>	Fungi
EDK04957	NCBI		M.gris	<i>Magnaporthe grisea 70-15</i>	Fungi
EAW17746	NCBI		N.fisc	<i>Neosartorya fischeri NRRL 181</i>	Fungi
EDN92841	NCBI		S.scle	<i>Sclerotinia sclerotiorum 1980</i>	Fungi
EAK86428	NCBI		U.mayd	<i>Ustilago maydis</i>	Fungi
CAX41987	NCBI		C.dub	<i>Candida dubliniensis CD36</i>	Fungi
XP_001244656	NCBI	http://genome.jgi-psf.org/Pospl1/Pospl1.download.ftp.html	C.imm	<i>Coccidioides immitis RS RS</i>	Fungi
BAH80450	NCBI		L.edod	<i>Lentinula edodes</i>	Fungi
56379		http://genome.jgi-psf.org/Pospl1/Pospl1.download.ftp.html	P.plac	<i>Postia placenta Mad-698</i>	Fungi
107376		http://genome.jgi-psf.org/Pospl1/Pospl1.download.ftp.html	P.plac	<i>Postia placenta Mad-698</i>	Fungi
EEP81937	NCBI		U.rees	<i>Uncinocarpus reesii 1704</i>	Fungi
XP_504084	NCBI		Y.lipo	<i>Yarrowia lipolytica CLIB122 W29</i>	Fungi
CAG79677	NCBI		Y.lipo	<i>Yarrowia lipolytica CLIB122</i>	Fungi
CAC42207	NCBI		A.med	<i>Amycolatopsis mediterranei U32</i>	Bact_Act
ACU36800	NCBI		A.mir	<i>Actinosynnema mirum DSM 43827</i>	Bact_Act
AAD32751	NCBI		S.lav	<i>Streptomyces lavendulae</i>	Bact_Act
EAL65201	NCBI		D.disc	<i>Dictyostelium discoideum</i>	Amoe
EAL65166	NCBI		D.disc	<i>Dictyostelium discoideum</i>	Amoe
88764		http://genome.jgi-psf.org/Dicpu1/	D.pur	<i>Dictyostelium purpureum</i>	Amoe
92439		http://genome.jgi-psf.org/Dicpu1/	D.pur	<i>Dictyostelium purpureum</i>	Amoe
37473		http://genome.jgi-psf.org/Dicpu1/	D.pur	<i>Dictyostelium purpureum</i>	Amoe

Dataset S07 List of sequences and species included in the CBM2 analysis.

Accession numbers, source database, associated module (where applicable), abbreviated and full species name are indicated for every sequence retained in the phylogenetic analysis of CBM2 modules. Sequences are sorted by taxonomic group then by species name.

Taxonomic groups are as follows: PPN: Plant-Parasitic Nematode, with the following suffixes: _RKN for root-knot nematode, _CN for cyst nematode, _Rad for Radopholinae, _Ang for Anguinidae, _Aph for Aphelenchoidea; Bact: Bacteria, with the following suffix: _Prot: Proteobacteria, _Firm: Firmicutes, _Act: Actinobacteria. Third column's associated modules abbreviations are as follows: CBM10: Carbohydrate-binding module, family 10; CBM33: Carbohydrate-binding module, family 33; CBP: no associated module, Cellulose-binding protein; CE1: Carbohydrate-esterase family1 catalytic module; EXPN: Expansin-like module; GH10: Glycoside hydrolase family 10 catalytic module; GH12: Glycoside hydrolase family 12 catalytic module; GH18: Glycoside hydrolase family 18 catalytic module; GH5: GH5 cellulase catalytic module; GH6: Glycoside hydrolase family 6 catalytic module.

Accession	Source	Assoc. Module	Species	Species long	Group
CAM33387	NCBI	CBP	M.are	<i>Meloidogyne arenaria</i>	PPN_RKN
CAM33384	NCBI	CBP	M.are	<i>Meloidogyne arenaria</i>	PPN_RKN
Mh_Contig344.4481..6245	http://www.hapla.org	GH5	M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Mh_Contig392.21227..22975	http://www.hapla.org	GH5	M.hap	<i>Meloidogyne hapla</i>	PPN_RKN
Minc09446a	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc09298a	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc13221a	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAD45868	NCBI	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAK21882	NCBI	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAR37374	NCBI	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc03286	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc14048	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc00168	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc09177	http://www.inra.fr/meloidogyne_incognita	EXPN	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc14116	http://www.inra.fr/meloidogyne_incognita	EXPN	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc15746	http://www.inra.fr/meloidogyne_incognita	CBP	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
CAM33389	NCBI	CBP	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
CAM33385	NCBI	CBP	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
CAM33388	NCBI	CBP	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
AAC05133	NCBI	CBP	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc14047a	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc00169a	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
Minc03287	http://www.inra.fr/meloidogyne_incognita	GH5	M.inc	<i>Meloidogyne incognita</i>	PPN_RKN
CAJ77137	NCBI	GH5	M.jav	<i>Meloidogyne javanica</i> HYN	PPN_RKN
CAM33392	NCBI	CBP	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN
CAM33390	NCBI	CBP	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN
CAM33391	NCBI	CBP	M.jav	<i>Meloidogyne javanica</i>	PPN_RKN
ABV54446	NCBI	GH5	R.sim	<i>Radopholus similis</i>	PPN_Rad
ABV54447	NCBI	GH5	R.sim	<i>Radopholus similis</i>	PPN_Rad

ABX79356	NCBI	GH5	P.cof	<i>Pratylenchus coffeae</i>	PPN_Prat
BAB68522	NCBI	GH5	P.pen	<i>Pratylenchus penetrans</i>	PPN_Prat
AAC48325	NCBI	GH5	G.ros	<i>Globodera rostochiensis RO-1</i>	PPN_CN
CAC83611	NCBI	EXPN	G.ros	<i>Globodera rostochiensis RO1</i>	PPN_CN
AAD56392	NCBI	GH5	G.tab	<i>Globodera tabacum solanacearum</i>	PPN_CN
AAC48327	NCBI	GH5	H.gly	<i>Heterodera glycines</i>	PPN_CN
AAN32887	NCBI	CBP	H.gly	<i>Heterodera glycines</i>	PPN_CN
CAC12958	NCBI	GH5	H.sch	<i>Heterodera schachtii</i>	PPN_CN
ABY49997	NCBI	CBP	H.sch	<i>Heterodera schachtii</i>	PPN_CN
ABY52965	NCBI	GH5	D.afr	<i>Ditylenchus africanus</i>	PPN_Ang
ACJ60676	NCBI	GH5	D.des	<i>Ditylenchus destructor</i>	PPN_Ang
ACE83181	NCBI	GH5	C.jap	<i>Cellvibrio japonicus Ueda107</i>	Bact_Prot
ACE85262	NCBI	CBP	C.jap	<i>Cellvibrio japonicus Ueda107</i>	Bact_Prot
CAA60493	NCBI	GH5-CBM10	C.jap	<i>Cellvibrio japonicus Ueda107</i>	Bact_Prot
ABD79328	NCBI	GH5	S.deg	<i>Saccharophagus degradans 2-40</i>	Bact_Prot
AAA23233	NCBI	GH5	C.cel	<i>Clostridium cellulovorans ATCC 35269</i>	Bact_Firm
ACU54908	NCBI	GH5-GH18	A.fer	<i>Acidimicrobium ferrooxidans DSM 10331</i>	Bact_Act
ACU37826	NCBI	CE1	A.mir	<i>Actinosynnema mirum DSM 43827</i>	Bact_Act
ACU37122	NCBI	GH10	A.mir	<i>Actinosynnema mirum DSM 43827</i>	Bact_Act
ABJ29232	NCBI	GH12	Act	<i>Actinomycetales</i>	Bact_Act
AAF00931	NCBI	GH18	Cell	<i>Cellulomonas sp. GM13</i>	Bact_Act
BAH26283	NCBI	CBP	M.bov	<i>Mycobacterium bovis BCG str. Tokyo 172</i>	Bact_Act
CAD96862	NCBI	CBP	M.bov	<i>Mycobacterium bovis AF2122/97</i>	Bact_Act
ACR15963	NCBI	CBP	M.can	<i>Mycobacterium canettii CIP140010059</i>	Bact_Act
ACC43206	NCBI	CBM2	M.mar	<i>Mycobacterium marinum M</i>	Bact_Act
ACT25069	NCBI	CBP	M.tub	<i>Mycobacterium tuberculosis KZN 1435</i>	Bact_Act
ABR06353	NCBI	CBP	M.tub	<i>Mycobacterium tuberculosis F11</i>	Bact_Act
ABQ73759	NCBI	CBP	M.tub	<i>Mycobacterium tuberculosis H37Ra</i>	Bact_Act
AAK46316	NCBI	CBP	M.tub	<i>Mycobacterium tuberculosis CDC1551</i>	Bact_Act
CAA98397	NCBI	CBP	M.tub	<i>Mycobacterium tuberculosis H37Rv</i>	Bact_Act
ABL03093	NCBI	CBM2	M.ulc	<i>Mycobacterium ulcerans Agy99</i>	Bact_Act
CBG75198	NCBI	CE1	R.sca	<i>Streptomyces scabiei 87.22</i>	Bact_Act
CBG67793	NCBI	GH12	R.sca	<i>Streptomyces scabiei 87.22</i>	Bact_Act
ABV98993	NCBI	GH18	S.are	<i>Salinispora arenicola CNS-205</i>	Bact_Act
BAC68265	NCBI	GH12	S.ave	<i>Streptomyces avermitilis MA-4680</i>	Bact_Act
BAC70589	NCBI	GH18	S.ave	<i>Streptomyces avermitilis MA-4680</i>	Bact_Act
BAC69965	NCBI	CBM33	S.ave	<i>Streptomyces avermitilis MA-4680</i>	Bact_Act
AAD16453	NCBI	GH18	S.coel	<i>Streptomyces coelicolor A3(2)</i>	Bact_Act
CAB61599	NCBI	GH12	S.coel	<i>Streptomyces coelicolor A3(2)</i>	Bact_Act
AAC45429	NCBI	GH12	S.hal	<i>Streptomyces halstedii JM8</i>	Bact_Act
BAF49409	NCBI	GH18	S.hal	<i>Streptomyces halstedii MF425</i>	Bact_Act
AAQ31831	NCBI	GH12	S.liv	<i>Streptomyces lividans TK23</i>	Bact_Act
AAB71950	NCBI	GH12	S.liv	<i>Streptomyces lividans 1326</i>	Bact_Act

CAA52139	NCBI	GH12	S.roc	<i>Streptomyces rochei</i> A2	Bact_Act
AAF91283	NCBI	GH12	S.sp.	<i>Streptomyces</i> sp. 11AG8	Bact_Act
BAE47186	NCBI	GH18	S.sp.	<i>Streptomyces</i> sp. AJ9463	Bact_Act
ABP55128	NCBI	GH5	S.trop	<i>Salinispora tropica</i> CNB-440	Bact_Act
AAD25090	NCBI	GH12	S.vir	<i>Streptomyces viridosporus</i> T7A	Bact_Act
ACN56471	NCBI	GH12	S.xyl	<i>Streptomyces xylophagus</i> KX6	Bact_Act
P26414	NCBI	GH6	T.bis	<i>Thermobispora bispora</i>	Bact_Act
AAC09379	NCBI	GH5	T.fus	<i>Thermobifida fusca</i> YX	Bact_Act