A Cellulose Loosening Protein is One of Nature's Most Widely Shared Tools
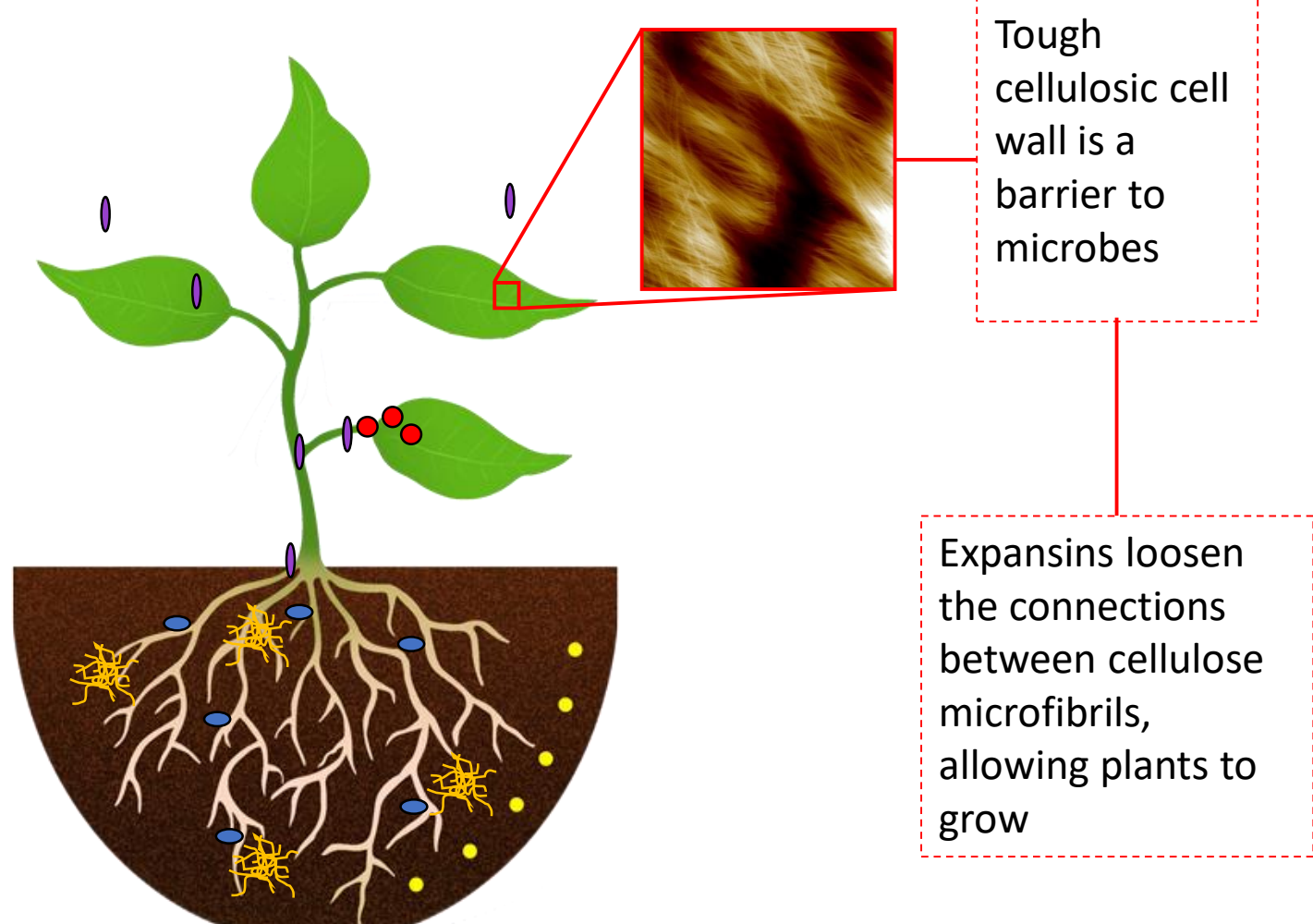
William R. Chase, Olga Zhaxybayeva, Rob Dunn, Daniel J. Cosgrove, Lori R. Shapiro

PennState · NC STATE UNIVERSITY · HARVARD UNIVERSITY · DARTMOUTH

Summary

Background

- Plants use proteins called expansins during normal growth and development to loosen and rearrange the complex network of polysaccharides that comprise their cell walls

- Expansin proteins were discovered in microbes, but their biological functions, significance, and evolutionary history remain understudied

- We use the recent expansion in publicly available genome data to survey the diversity, evolutionary history, and ecological significance of expansin proteins in microbes

- Plants and microbes live in complex communities

- Plant-microbe association ranges from symbiotic to pathogenic

- Experiments show that expansin is a virulence factor in some microbes, and promotes beneficial plant-microbe association in others

Tough cellulosic cell wall is a barrier to microbes

Expansins loosen the connections between cellulose microfibrils, allowing plants to grow

Methods

- Public databases were queried for the presence of expansin by BLAST and keyword annotation search

- Sequences were filtered, aligned, and trimmed. Then the best fitting model of evolution was inferred using ModelFinder

- Phylogenetic trees were built using maximum likelihood (IQ-TREE) and Bayesian (MrBayes) methods

- HGT events were inferred by the placement of a taxa in the gene tree outside of its expected species grouping; these putative HGT events were then validated by examining concordance and node support across several methods of phylogenetic inference

- Evidence of ancient and recent HGT events was inferred by finding the most parsimonious explanation for the present day distribution of expansins across the tree of life using a combination of gene duplication, gene transfer, and gene loss (duplication-transfer-loss modeling)

- A literature review was performed to classify each expansin-containing organism into a general ecological niche

Key Findings

- Expansin originated in ancient time before the rise of land plants (between 1.5 – 0.8 Ga, origin lineage is unclear)

- From there, expansin was spread across the tree of life by vertical and horizontal transfer between and within domains

- Expansin was adapted for two main purposes: to loosen structural cellulose, and to aid plant-microbe interactions

- Expansin appears to be required by organisms that use cellulose structurally, and as an accessory gene in other organisms

- Expansin is found in a highly diverse set of taxa, is adapted to various functions spanning multiple ecological niches, and can fuse with several other carbohydrate-active modules for form novel chimeric proteins

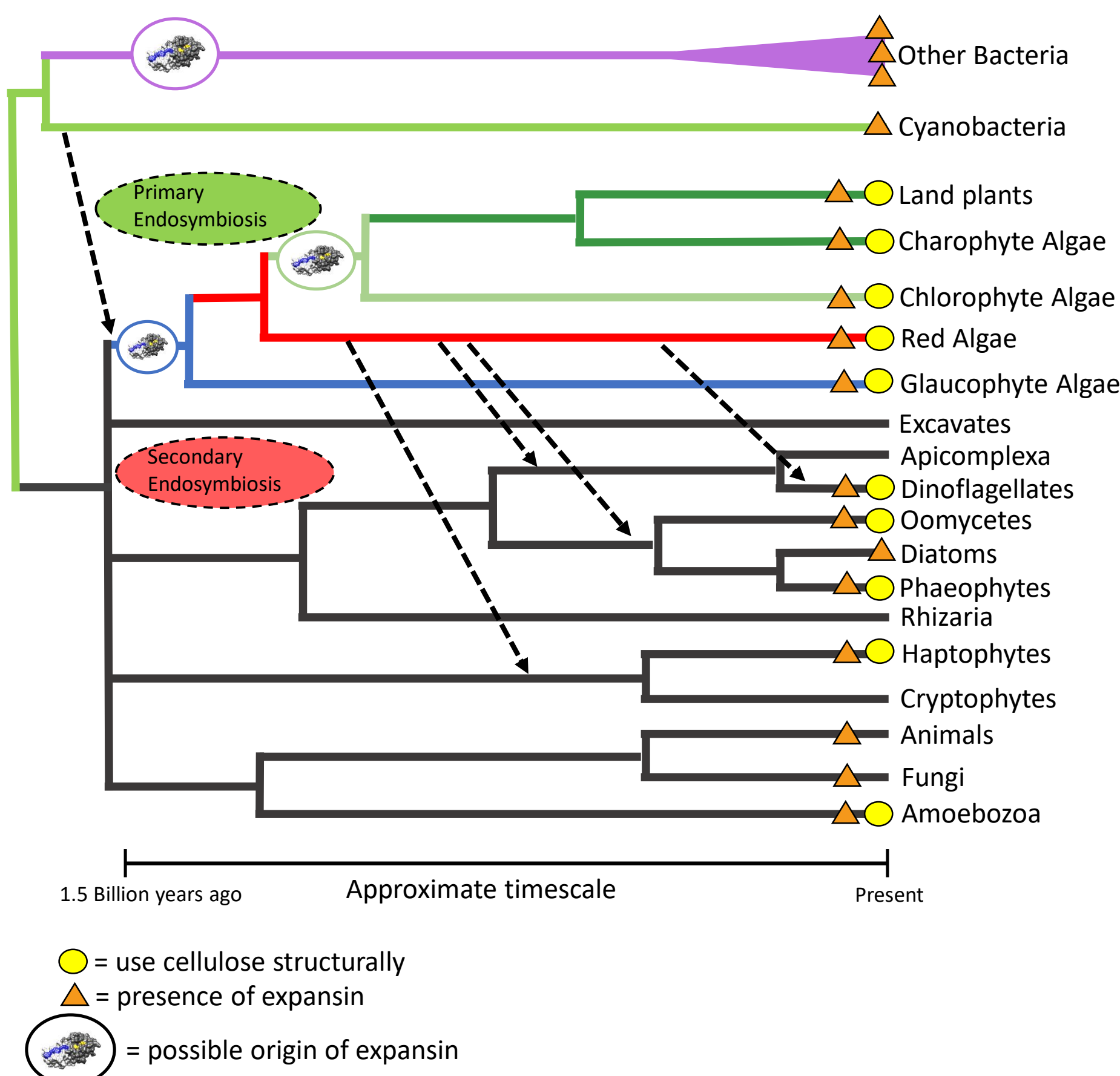## Expansins are widespread across diverse taxa

Figure 1: A cartoon representation of the tree of life with expansin-containing lineages and cellulose-utilizing lineages are denoted. Branch length is not to scale.

- = use cellulose structurally
- = presence of expansin
- = possible origin of expansin

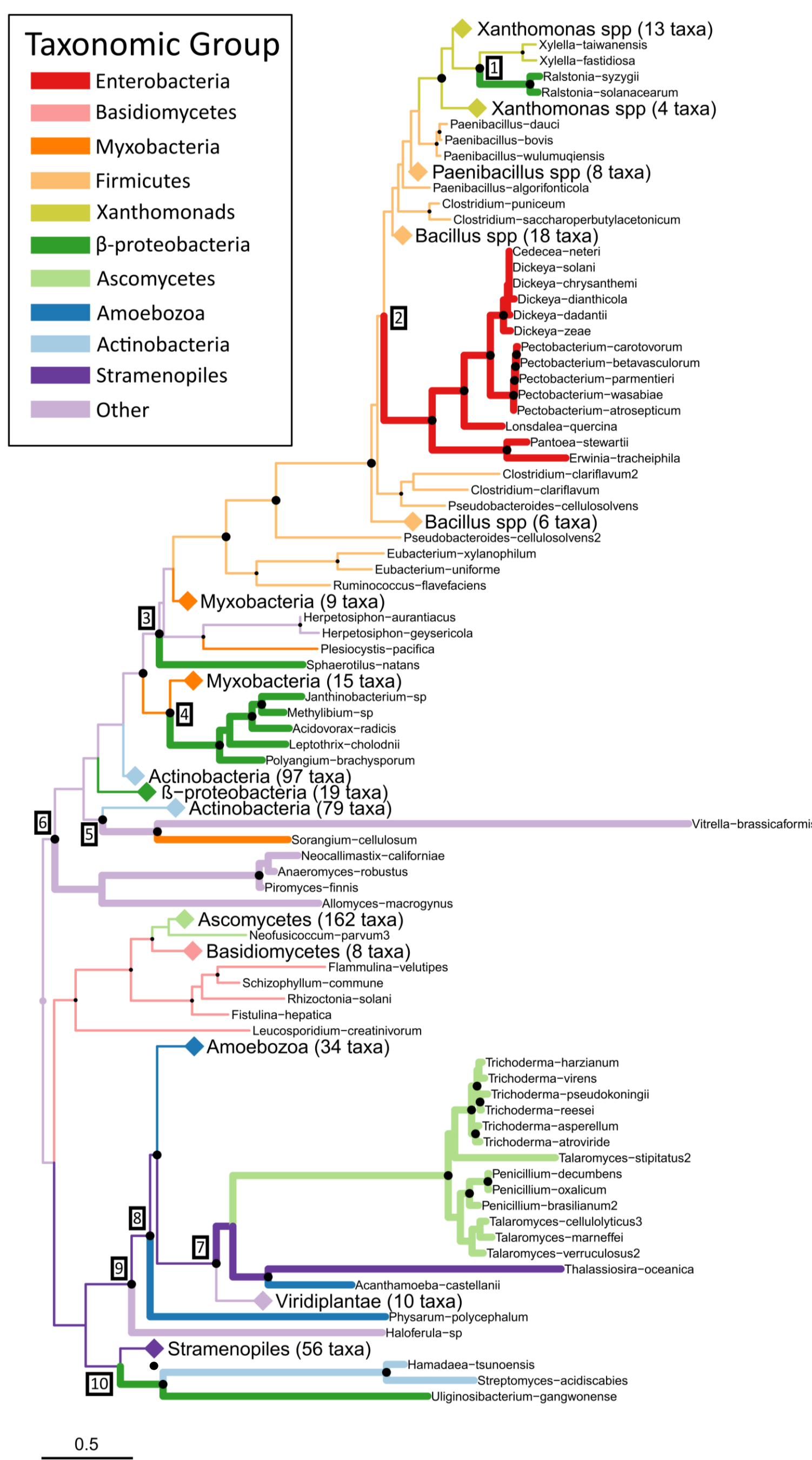## Expansin evolution was shaped by horizontal gene transfer

Figure 2: Maximum likelihood gene tree of all microbial expansin proteins. Larger branches indicate putative HGT recipients. Well supported nodes are marked with black circles. Scale bar represents amino acid substitutions per site.

Table 1: Summary of the 10 putative HGT events identified in the original expansin gene tree. Support at each HGT node was evaluated by multiple phylogenetic analyses. Nodes are highlighted according to confidence based on support values. Blue nodes are very high certainty, yellow are well-supported, and white have questionable support.

Node Number	HGT recipient	HGT recipient groups within	Full ML tree support (SH-aLRT/UFboot)	Full bayesian tree support (posterior probability)	Pruned ML tree support (SH-aLRT/UFboot)
1	Ralstonia spp. (β-proteobacteria)	Xanthomonads	97.9/100	100	98/100
2	Enterobacteria *	Firmicutes	44.3/85	polytomy	21.9/84
3	Sphaerotilus natans (β-proteobacteria)	Myxobacteria	89.6/99	polytomy	93.3/93
4	Janthinobacterium sp., Methylibium sp., Acidovorax radicis, Leptothrix cholodnii, Polyangium brachysporum (β-proteobacteria)	Myxobacteria	93.5/99	97	92.2/97
5	Vitrella brassicaformis (Alveolate), Sorangium cellulosum (β-proteobacteria)	Actinobacteria	87.4/99	52	83.8/89
6	Neocallimastix californiae, Anaeromyces robustus, Piromyces finnis, Allomyces macrogynus (Chytrid anaerobic gut fungi)	Actinobacteria	87.2/98	76	93.7/89
7	Fungal expansin-CBM1 fusion proteins (Ascomycetes)	Phytophthora	91.5/93	100	95.1/99
8	Physarum polycephalum (Amoebozoa)	Phytophthora	96/98	76	N/A
9	Haloferula sp. (Verrucomicrobia)	Phytophthora	85/97	100	95.9/100
10	Hamadaea tsuonensis (Actinobacteria), Streptomyces acidiscabies (Actinobacteria), Uliginosibacterium gangwonense (β-proteobacteria)	Amoebozoa	97.2/98	79	84.7/86

= interdomain HGT event

* = low bootstrap support, but recovered across 12 of 13 tree-building runs and supported by a DTL model

Table 2: Approximate age of putative HGT events inferred by parsimony.

Putative HGT acceptor	Estimated timescale	Within-group Sequence Identity
MRCA of Myxococcales	1,236 MYA	0.538
MRCA of Micromonospora	198 MYA	0.803
MRCA of Nocardia	125 MYA	0.594
MRCA of Xanthomonas	30 MYA	0.784
MRCA of Xylella	<0.936 MYA	0.941
MRCA of Lentzea	0.425 MYA	0.767
MRCA of Herbidospora	0.0315 MYA	0.911
MRCA of Pectobacterium	0.0174 MYA	0.992
MRCA of Dickeya	0.0058 MYA	0.97
Ralstonia solanacearum	Recent	-
Clavibacter michiganensis	Recent	-
Clostridium puniceum	Recent	-
Erwinia tracheiphila	Recent	-
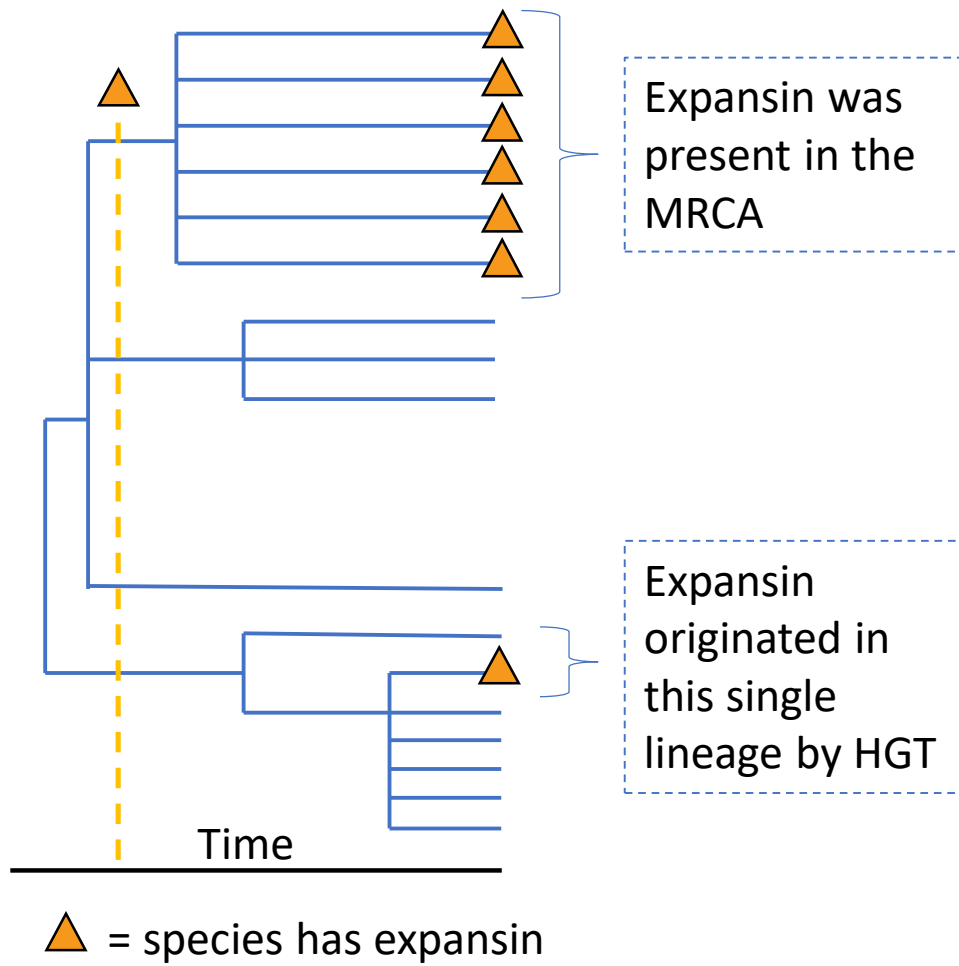
Figure 3: A simplified version of the parsimony model used to infer HGT in table 2.

Species Tree — Expansin was present in the MRCA / Expansin originated in this single lineage by HGT

= species has expansin

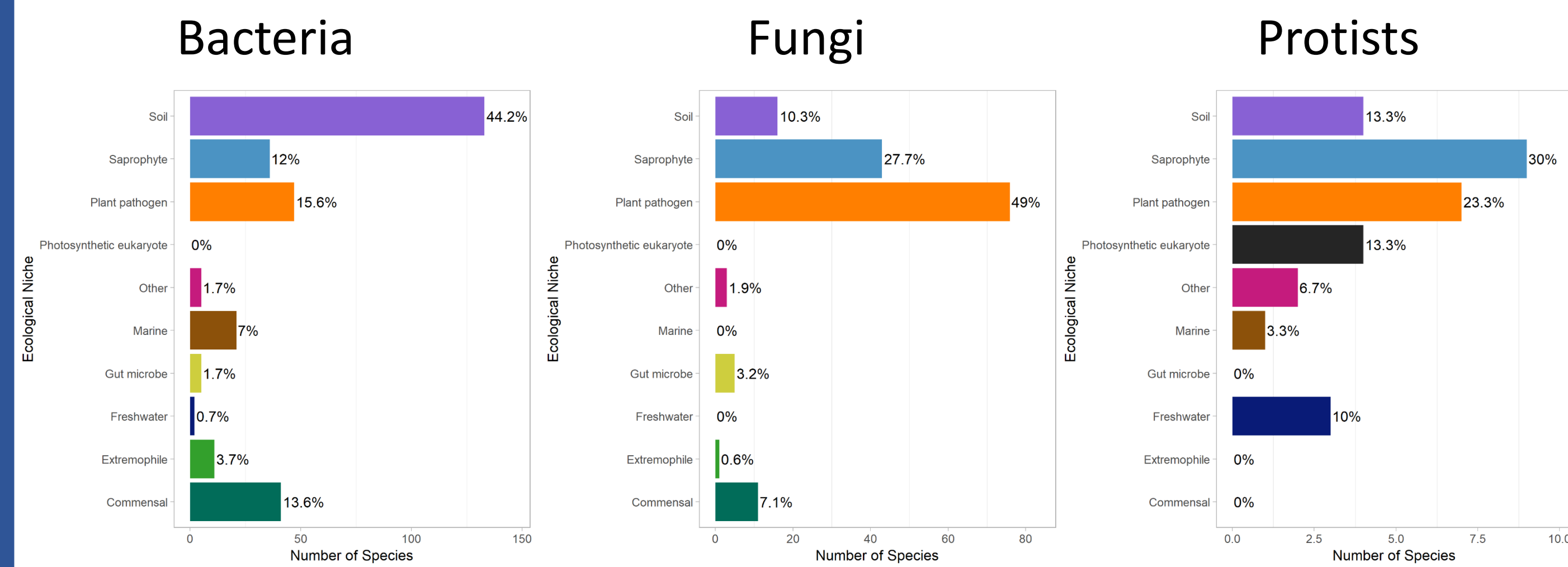## Expansins are found in numerous ecological niches

Figure 4: Ecological niches of expansin-containing microbes from bacteria, fungi, and protists.

Expansins are abundant in the world's worst plant pathogens

Table 3: Although most expansin-containing microbes are non-pathogenic, expansins are overrepresented among the worlds most damaging agricultural pathogens. The "top 10" worst bacterial and fungal plant pathogens are listed, and those with expansin are noted by the accession number of the expansin gene.

Bacteria	Accession Number	Fungi	Accession Number
Pseudomonas syringae	—	Magnaporthe oryzae	XP_003711472
Ralstonia solanacearum	AOE88952	Botrytis cinerea	CCD54783
Agrobacterium tumefaciens	—	Fusarium graminearum	CEF78987
Xanthomonas oryzae	WP_011409499	Fusarium oxysporum	KNB11838
Xanthomonas campestris	WP_011038627	Blumeria graminis	—
Xanthomonas axonopodis	WP_048490938	Mycosphaerella graminicola†	EGP90221
Erwinia amylovora	—	Ustilago maydis	—
Xylella fastidiosa	SHG20409	Melampsora lini‡	—
Dickeya (dadantii and solani)	WP_038924663, KHN53802	Puccinia spp.	KNZ59084, EFP85064, KNF04021
Pectobacterium (carotovorum and atrosepticum)	SHG09776, AIK14148	Colletotrichum spp.	OLN88490, CCF37207, KXH45263, XP_007594649, KXH43193, KXH26982, ENH88520, EQB51143, KDN66674

† Reclassified as Zymoseptoria tritici
‡ No available genome sequence

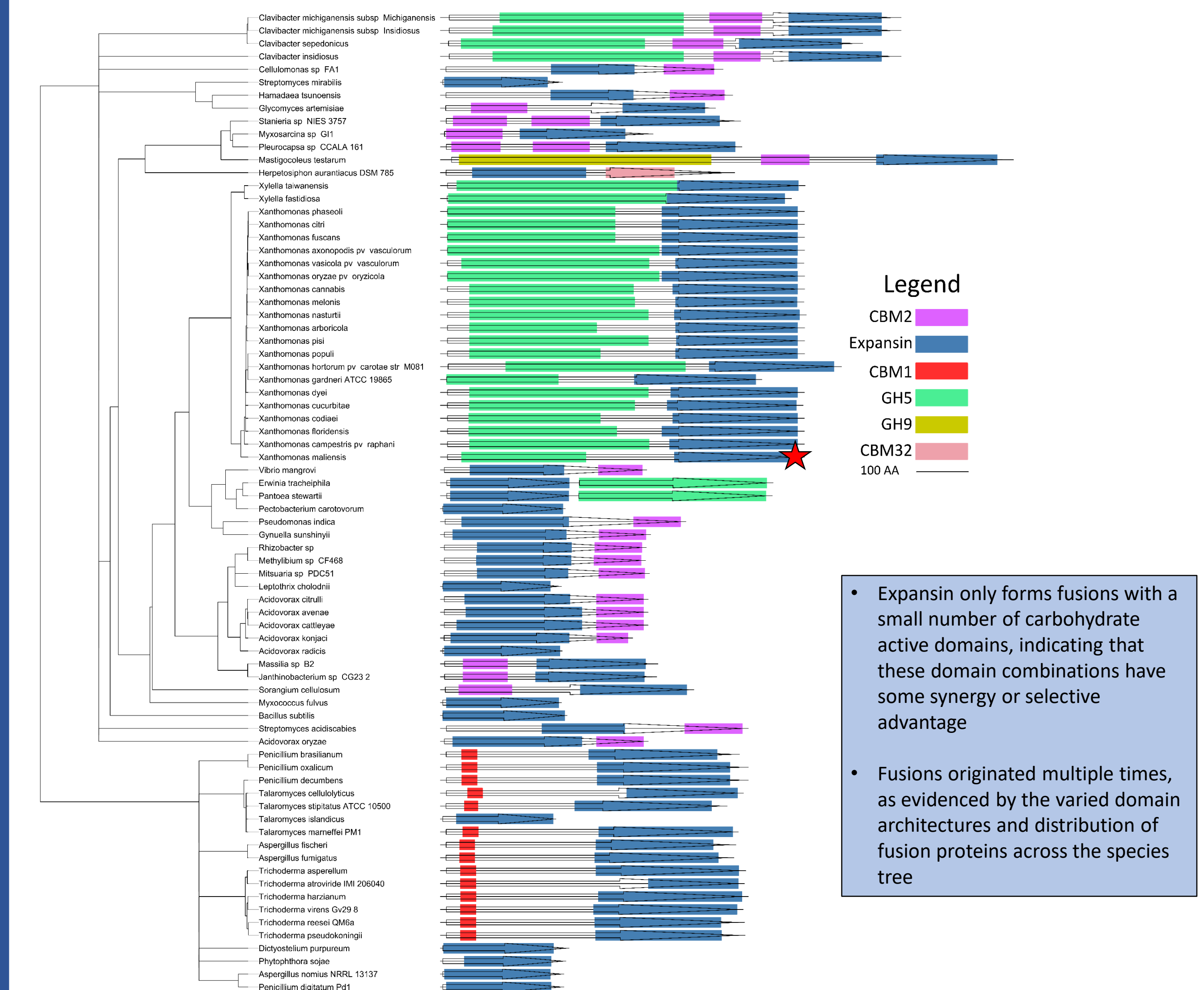## Expansins are actively innovating through gene fusion

Figure 5: The domain architecture of expansin-fusion proteins and selected non-fusion proteins. Frequent fusion modules are glycoside hydrolase domains (GH) and carbohydrate binding modules (CBM). The phylogenetic tree is the species tree built from conserved proteins gyrB for bacteria and EFI-a for fungi. The red star denotes two lineages where the expansin and associated domain are separate ORFs, but adjacent in the genome and separated by only a few nucleotides.

Legend: CBM2, Expansin, CBM1, GH5, GH9, CBM32, 100 AA

- Expansin only forms fusions with a small number of carbohydrate active domains, indicating that these domain combinations have some synergy or selective advantage

- Fusions originated multiple times, as evidenced by the varied domain architectures and distribution of fusion proteins across the species tree

References

- Cosgrove, D. J. (2017). Microbial expansins. Annual review of microbiology, 71, 479-497.
- Nikolaidis, N., Doran, N., & Cosgrove, D. J. (2013). Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. Molecular biology and evolution, 31(2), 376-386.
- Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. Nature, 480(7376), 241.