

DR LORI R. SHAPIRO (Orcid ID : 0000-0002-8794-6485)

Article type : - Regular Manuscript

Global cellulose biomass, horizontal gene transfers and domain fusions drive microbial expansin evolution

William R. Chase¹

Olga Zhaxybayeva^{2,3}

Jorge Rocha⁴ ORCID 0000-0002-8083-4325

Daniel J. Cosgrove¹

Lori R. Shapiro^{4, §} ORCID 0000-0002-8794-6485

Corresponding author: Lori Shapiro (Lori.R.Shapiro@gmail.com), 770-713-1871

¹Department of Biology, Pennsylvania State University, University Park, PA, USA, 16801

²Department of Biological Sciences, Dartmouth College, Hanover, NH, USA 03755

³Department of Computer Science, Dartmouth College, Hanover, NH, USA 03755

⁴Department of Microbiology and Immunology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston MA, USA 02115

§ Current Address: Department of Applied Ecology, North Carolina State University, 100 Brooks Avenue, Raleigh, NC, USA 27695

Received: 21 June 2019

Accepted: 19 December 2019

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/nph.16428](https://doi.org/10.1111/nph.16428)

This article is protected by copyright. All rights reserved

Keywords: expansin, cellulose, horizontal gene transfer, domain fusion, cell wall, plant-microbe interactions, glycoside hydrolase, carbohydrate binding module.

Author Contributions: LRS conceived of the study, LRS, OZ and WRC designed analysis, WRC and LRS performed analyses; WRC, OZ, DJC, JR, LRS critically interpreted analyses, LRS wrote the first draft of the manuscript; WRC, OZ, DJC, JR, LRS contributed manuscript revisions.

Data Deposition Statement: No new data were generated for this study. The data gathered from NCBI, and the custom scripts created to analyze these data, are available at https://github.com/will-r-chase/microbe_expansin.

Abstract

- Plants must rearrange the network of complex carbohydrates in their cell walls during normal growth and development. To accomplish this, all plants depend on proteins called expansins that non-enzymatically loosen noncovalent bonding between cellulose microfibrils.
- Surprisingly, expansin genes have more recently been found in some bacteria and microbial eukaryotes, where their biological functions are largely unknown.
- Here, we reconstruct a comprehensive phylogeny of microbial expansin genes. We find these genes in all eukaryotic microorganisms that have structural cell wall cellulose, suggesting expansins evolved in ancient marine microorganisms long before the evolution of land plants. We also find expansins in an unexpectedly high diversity of bacteria and fungi that do not have cellulosic cell walls. These bacteria and fungi inhabit varied ecological contexts mirroring the diversity of terrestrial and aquatic niches where plant and/or algal cellulosic cell walls are present.
- The microbial expansin phylogeny shows evidence of multiple horizontal gene transfer events within and between bacterial and eukaryotic microbial lineages, which may in part underlie their unusually broad phylogenetic distribution. Overall, expansins are unexpectedly

widespread in bacteria and eukaryotes, and the contribution of these genes to microbial ecological interactions with plants and algae has likely been underappreciated.

Introduction

Cellulose – a linear polysaccharide comprised of hundreds to thousands of D-glucose units – is the most abundant biopolymer on Earth. The vast majority of global cellulose biomass is present in plant cell walls, where cellulose microfibrils are interlinked with hemicelluloses and pectins to provide structural support (Cosgrove, 2000; Li *et al.*, 2003; Bar-On *et al.*, 2018). This carbohydrate and protein matrix allows plant cell walls to withstand high tensile stresses, which can reach as high as 1000 atmospheres during growth (Cosgrove, 2005). The strength of the structural carbohydrates in the cell wall also creates a formidable physical barrier against pathogenic microorganisms (Hématy *et al.*, 2009). Cellulosic cell walls similar in structure to those in plants also occur in some algal and other microbial eukaryotic groups (Darley *et al.*, 2003; Popper & Tuohy, 2010; Popper *et al.*, 2011). Tunicates (Urochordata) are the only metazoan group known to use cellulose structurally, and are thought to have acquired cellulose synthase genes horizontally from bacteria (Kimura & Itoh, 1995; Nakashima *et al.*, 2004; Tamai *et al.*, 2004).

All of these diverse organisms are confronted with the dilemma of how to loosen their cellulose-based matrix of structural carbohydrates in order to expand their cell walls during normal growth and development. In plants and some green algae, non-enzymatic proteins called expansins provide the most important functions for loosening structural cellulose, and expansins are most highly expressed during active growth in any tissue where cell wall extension is critical (Brummell *et al.*, 1999; Chen & Bradford, 2000; Cho & Cosgrove, 2000; Im *et al.*, 2000; Pien *et al.*, 2001; Lee *et al.*, 2003; Gray-Mitsumune *et al.*, 2004; Cosgrove, 2016; Ramakrishna *et al.*, 2019). Expansin proteins are tightly packed two-domain structures of 200-250 amino acids with a planar polysaccharide binding surface (Supplemental Figure 1). The N-terminal domain is related to family 45 glycoside hydrolases, but lacks lytic activity. The C-terminal domain is related to group 2 grass pollen allergens (Cosgrove, 2000; Kende *et al.*, 2004; Sampedro & Cosgrove, 2005; Yennawar *et al.*, 2006a; Kerff *et al.*, 2008b; Lombard *et al.*, 2013).

Expansin genes are universally present in all plant species, and most plant genomes contain multiple expansin homologs (Li *et al.*, 2002; Ding *et al.*, 2016; Santiago *et al.*, 2018). In vascular plants, expansins have diversified from a common ancestor into four distinct genetic subfamilies. Two of these subfamilies, the α and β expansins, have been empirically shown to cause irreversible cell wall extension (Cosgrove, 2000; Lee *et al.*, 2001; Li *et al.*, 2002; Li *et al.*, 2003; Cosgrove, 2015). The gene sequences of the two remaining subfamilies, expansin-like families A and B (EXLA and EXLB, respectively), contain both canonical expansin domains but neither EXLA nor EXLB have yet been functionally characterized. The current working hypotheses for expansin mode of action is disruption of noncovalent bonds at biomechanical hotspots between cellulose microfibrils, or between cellulose microfibrils and hemicellulose. Disruption of such bonds allows slippage of carbohydrate polymers at load bearing elements of the cell wall. This causes water uptake and irreversible enlargement of the cell without compromising tensile strength (Wang *et al.*, 2013; Cosgrove, 2015; Cosgrove, 2016). The ubiquity of expansins in land plants and some green algae, the phylogenetic diversity of expansins in vascular plants, and their essential role in cell wall growth underlies the hypotheses that expansins may have first evolved in green algae and then diversified in land plants, and that these genes were necessary for the evolutionary success and adaptive radiation of the Plantae lineage (Schipper *et al.*, 2002; Carey & Cosgrove, 2007; Carey *et al.*, 2013; Nikolaidis *et al.*, 2013; Cosgrove, 2017).

What has long remained unknown is the distribution and function of expansin genes in non-Plantae organisms – especially those that do not have cellulosic cell walls (Brotman *et al.*, 2008; Ogasawara *et al.*, 2009; Nikolaidis *et al.*, 2013; Georgelis *et al.*, 2015; Cosgrove, 2017). Fungal and bacterial genes predicted to have similar structure as β -expansins were first identified once databases began accumulating large numbers of genomic sequences. These microbial expansin-like genes were assigned to a newly established EXLX subfamily (Kende *et al.*, 2004; Kerff *et al.*, 2008b). A single copy expansin gene (Exlx1) from the rhizosphere plant commensal bacterium, *Bacillus subtilis*, was the first bacterial expansin gene with the predicted two domain structure of a plant expansin to be identified. This gene, referred to by expansin naming convention as BsExlx1, is highly divergent from plant expansins at the amino acid sequence level, but contains a conserved aspartic acid in domain 1

that is crucial for cell wall extension, and linear aromatic residues in domain 2 that are essential for polysaccharide binding (Supplemental Figure 1) (Kerff *et al.*, 2008b; Georgelis *et al.*, 2015).

Bacillus subtilis, like many species of bacteria, utilizes cellulose as part of an extracellular biofilm matrix, but does not use cellulose as a structural component of their cell walls (Schleifer & Kandler, 1972; Branda *et al.*, 2005). Functional characterization of BsExlx1 found this gene significantly increases the efficiency of epiphytic maize root colonization by *B. subtilis*, despite showing 10 times less *in vitro* cell wall loosening activity than plant expansins (Kerff *et al.*, 2008b). This suggests that the function of expansins in microbial backgrounds may be to facilitate ecological interactions with, or colonization of hosts that produce cellulosic cell walls (Meibom *et al.*, 2004; Burke *et al.*, 2011; Egan *et al.*, 2013; Haney *et al.*, 2015; Shabat *et al.*, 2016; Niu *et al.*, 2017). Since this first characterization of an expansin gene in a bacterium, additional EXLX family expansin genes have been identified in phylogenetically diverse bacteria, fungi and some microbial eukaryotes (reviewed in (Nikolaidis *et al.*, 2013; Cosgrove, 2017)). Like plant expansins, no microbial expansins have been documented to have enzymatic activity (Jahr *et al.*, 2000; Laine *et al.*, 2000; Olarte-Lozano *et al.*, 2014; Georgelis *et al.*, 2015; Cosgrove, 2017; Tancos *et al.*, 2018). Microbial expansin evolutionary history, taxonomic distribution, mechanism(s) of action, and ecological function(s) remains enigmatic in non-Plantae genetic backgrounds, and there is currently no framework for predicting their functional roles (Jahr *et al.*, 2000; Saloheimo *et al.*, 2002; Brotman *et al.*, 2008; Georgelis *et al.*, 2015; Junior *et al.*, 2015; Hwang *et al.*, 2019).

Here, we examine the phylogenetic distribution of expansin genes, and infer their possible ecological roles through four steps. First, we searched for expansin genes in all non-Plantae records of GenBank, and reconstructed the phylogeny of microbial expansin homologs in the context of the broader tree of life. Second, we considered the distribution of expansin genes relative to the life history of microorganisms that have them. Third, we analyzed the microbial expansin phylogeny for signals of horizontal gene transfer and hypothesized how ecological factors may be driving transfer of this gene between distantly related microbial taxa. Finally, we examined ongoing evolution of microbial expansin genes through fusions with carbohydrate active proteins. Together, these analyses indicate that microbial expansins are more widely utilized by microorganisms than previously

recognized, and that their distribution is in part driven by the underestimated importance of microbial expansins for mediating bacterial and fungal interactions with live and dead plant and algal matter.

Materials and Methods

Detection of microbial expansin sequences

Amino acid sequences encoding microbial expansins were identified in a two-step approach. In the first step, the NCBI non-redundant (*nr*) protein sequence database was queried using the keywords ‘expansin’ and ‘swollenin’ and excluding hits from Viridiplantae taxa (accessed Jan. 2017). Both ‘expansin’ and ‘swollenin’ were used as keywords because expansin sequences are submitted with both names to the database. This initial search returned ~11,500 amino acid sequences, which were curated to remove duplicates and ensure that all hits were *bona fide* microbial expansin genes. All hits were evaluated based on presence of the canonical expansin domains and key amino acid motifs (conserved aspartic acid in domain 1 and conserved aromatic triplet in domain 2) to the experimentally validated expansin sequences BsEXLX1 from *Bacillus subtilis* (AAB84448.1) and the alpha expansin AtEXPA4 from *Arabidopsis* (AEC09708.1). BsEXLX1 and AtEXPA4 were used as references because they represent the microbial expansin (BsEXLX1) or plant expansin (AtEXPA4) superfamilies, and the expansin function of both genes has been experimentally validated. Every sequence was searched for the presence of characteristic expansin domains using CD-Search (Marchler-Bauer *et al.*, 2016). Records were removed from the dataset if the amino acid sequence lacked either of the characteristic expansin domains or key residues (after (Kerff *et al.*, 2008b)). Amino acids sequences that flanked the two canonical expansin domains that were included in the GenBank record because of mis-annotation (such as RNA polymerase sequences) were trimmed to retain only the canonical expansin domains (and carbohydrate associated domains, when present). This procedure resulted in ~400 unique microbial expansin amino acid sequences.

The expansin sequences retrieved in this initial, keyword-based search were then manually separated into bacteria, fungi, and microbial eukaryotic subsets. Representative sequences from these three taxonomic groups were then used as BLASTP (version 2.6.0 accessed March 2017) (Altschul *et al.*, 1990) queries to identify any microbial expansin gene sequences in the *nr* database that may have been missed in the keyword search. Matches from Viridiplantae were excluded. This sequence-based

approach yielded ~200 additional hits that were added to the above described existing homolog list of ~400 unique microbial expansins. The sequences were again aligned to the reference expansin sequences (bacterial BsEXLX1 AAB84448.1, and plant AtEXPA4 AEC09708.1) using MAFFT v.7 with the FFT-NS-i option (maximum 1000 iterations) (Kato & Standley, 2013), and then manually filtered and trimmed to remove sequences that lack the canonical expansin GH45-like and CBM63 domains, and mis-annotated flanking regions. The final microbial expansin gene set contains 600 unique, *bona fide* non-Viridiplantae expansin proteins from 491 distinct microbial species (Notes S1). For 113 microbial species, there were at least two, and up to eight, non-identical expansin genes within the same species (Supplemental Table 1).

For all 491 microbial species taxonomy information was retrieved from the NCBI taxonomy database (Federhen, 2011). The presence of expansin genes was mapped onto the currently accepted phylogenies for Eukaryota and Bacteria (Federhen, 2011; Adl *et al.*, 2012; Hug *et al.*, 2016; Adl *et al.*, 2019). For each bacterial order with multiple species that contain expansin genes, the NCBI taxonomy database (Federhen, 2011) was used to determine the total number of named species and the total number of sequenced species.

Phylogenetic reconstruction

Because of high amino acid divergence, the expansin homologs from bacteria, fungi and the other eukaryotic microbes were aligned separately using MAFFT (option E-INS-i). Poorly aligned regions at the termini were manually trimmed to the alignment site that was shared across 90% of species. All three trimmed alignments were then combined, and re-aligned with MAFFT (option E-INS-i). Viridiplantae expansin gene sequences from one dicot (AtEXPA4, GenBank: O48818.1), one monocot (ZmEXPB1, GenBank: P58738.2), one non-vascular plant (PpEXPA10, XP_024392378.1), four sequences from the charophyte green algae *Klebsormidium nitens* (GAQ91800.1, GAQ85527.1, GAQ79710.1, GAQ91109.1) and two sequences from the chlorophyte green algae *Chlorella variabilis* (XP_005846210.1, XP_005846208.1) were added to the dataset. The dataset was re-aligned for a final time with MAFFT (option E-INS-i). The final sequence alignment contains 608 expansin sequences (600 microbial expansin sequences and 8 Viridiplantae reference sequences) and 689 amino acid sites (Notes S1).

The WAG+R7 model was chosen as the best evolutionary model based on the Akaike information criterion by ModelFinder (Kalyaanamoorthy *et al.*, 2017) (as implemented within IQ-tree v. 1.6; (Nguyen *et al.*, 2014)). The phylogeny was reconstructed in IQ-tree v. 1.6, using a smaller perturbation strength and larger number of stop iterations (options -pers 0.2 -nstop 500) to avoid local minima (all other parameters default). Node supports were estimated using the Shimodaira–Hasegawa like approximate likelihood ratio test (SH-aLRT) (Guindon *et al.*, 2010) and the ultrafast bootstrap with 1000 bootstrap pseudo-replicates (Hoang *et al.*, 2017). IQ-tree was run with these parameters 13 independent times to test the robustness of phylogenetic relationships. The resulting consensus tree of the 13 runs was manually rooted between prokaryotes and eukaryotes solely for presentation purposes (Notes S2). The tree was visualized and annotated using the ggtree package in R (v. 3.4.2) (R, 2015).

Inference of horizontal gene transfer events

Twenty-one putative horizontal gene transfer events were identified by finding incongruences between the expansin gene tree (Supplemental Figure 2) and the organismal taxonomy (Supplemental Figure 3). To further evaluate the strength of the relationships recovered at these 21 nodes, a Bayesian tree was reconstructed using MrBayes v. 3.2.6 (Ronquist *et al.*, 2012) on the XSEDE cluster (Towns *et al.*, 2014) (Supplemental Figure 4, Notes S1; full parameters are in Notes S5). Additionally, the full dataset of 600 microbial sequences was pruned to 350 representative taxa (Supplemental Figure 5, Notes S3, Notes S4). A maximum likelihood tree on the reduced dataset was reconstructed in IQ-tree with the same run parameters as the full tree (see above) using the WAG+G4 model, which was identified as the best substitution model by ModelFinder implemented within IQ-tree v. 1.6. For the Bayesian reconstruction, the expansins from *Vitrella brassicaformis* and *Emiliania huxleyi* were removed due to their extremely divergent sequences, to avoid problems with convergence. Bayesian trees were reconstructed in MrBayes v 3.2.6 (Ronquist *et al.*, 2012) on the XSEDE cluster (Towns *et al.*, 2014). Two independent runs were performed for 10 million generations, each with six chains using Metropolis-coupling with a heating parameter of 0.005 and swap frequency of 1. Each chain was sampled every 500 generations and the first 1.5 million samples were discarded as burn-in. The log likelihood of both runs plateaued after ~1.5 million generations (Supplemental Figure 6) and both runs converged on a similar tree (standard deviation of split frequencies between runs = 0.020738).

All parameters of the MCMC algorithm are listed in Notes S5. Because both runs converged on a similar tree, a majority rule consensus tree was constructed from the sampled trees of run 1 (Notes S6). For the maximum likelihood trees, node support was measured using the Ultrafast Bootstrap (UFboot) and the Shimdaira-Hasegawa-alternate likelihood ratio test (Guindon *et al.*, 2010; Minh *et al.*, 2013). For the Bayesian tree, the node support represents the posterior probability of the consensus tree.

Ecological niche determination and phylogenetic tree annotation

For each microbial species with an expansin gene, a literature search was carried out to determine the known ecological associations (Supplemental Table 4). For many species, there was little or no documentation of the life history. Furthermore, for many species, the existing descriptions of ecological life history may be incomplete. For example, the classifications of ‘plant commensal’, ‘saprophyte’, and ‘soil dweller’ likely share significant functional overlap, and many microbes may fit multiple of these overlapping categories. Many microbes thought of as ‘soil dwellers’ are likely also saprophytes, plant commensals, and/or plant pathogens depending on the environmental conditions (Carroll, 1988; Lofgren *et al.*, 2018; Selosse *et al.*, 2018). Despite these caveats, each species was assigned to only one of the following ecological categories after evaluating the available ecological information: freshwater, marine, gut microbe, soil dweller, plant commensal, plant pathogen, saprophyte, hot spring, sulfur mat, or wastewater. The certainty (or lack thereof) for the ecological assignments for each species is noted in Supplemental Table 4. The expansin microbial phylogeny was annotated with the collected ecological information in Supplemental Table 4 using the ggtree package in R (v. 3.4.2) (Yu *et al.*, 2017).

Identification of carbohydrate active domains fused to microbial expansin domains

A comprehensive list of microbial expansin genes fused to carbohydrate active domains was compiled by a search of the NCBI protein database (accessed April 2017) with the keywords “expansin” and “swollenin”. This was followed by a BLASTP search (Altschul *et al.*, 1990) with the expansin-swollenin fusion from *Trichoderma reesei* (Accession number: CAB92328.1) as a query. Both searches were constrained to records with a bit score above 100 and at least 300 amino acid residues

in length to exclude non-fused expansin genes, which are typically ~200-250 amino acids in length. The matches that met these two criteria were retained as putative genes with expansin-carbohydrate fusions.

The presence of a carbohydrate active domain(s) was then evaluated with a batch CD-search (Marchler-Bauer & Bryant, 2004) and DBSCAN search (Yin *et al.*, 2012). Records that shared more than 95% sequence identity to another record in the same species were considered redundant and were removed. The expansin domains were then aligned in MAFFT v7 (E-INS-i) with the plant (AtEXPA4) and bacteria (BsEXLX1) reference expansin sequences (Kato & Standley, 2013). The expansin - carbohydrate active domain fusion constructs were plotted next to the expansin gene tree using ggtree and genoPlotR (Guy *et al.*, 2010).

Plant and microbial expansin protein structures

The crystal structures of the bacterial expansin from *Bacillus subtilis* (BsEXLX1, PDB: 3D30) alone (Kerff *et al.*, 2008a) and in complex with plant cellobiose (PDB: 4FER) (Georgelis *et al.*, 2012), and the plant β -expansin ZmEXPB1 (PDB: 2HCZ) from *Zea mays* (Yennawar *et al.*, 2006b) were downloaded from the Protein Data Bank (PDB) (Rose *et al.*, 2017). The 3D protein structures were visualized with UCSF Chimera v1.2.2 (Pettersen *et al.*, 2004).

Results

Phylogenetic distribution of expansin genes across the tree of life

A survey of the NCBI *nr* database revealed a total of 600 unique expansin sequences in 491 species, in addition to those known in green algae (Chloroplastida) and terrestrial plants (Embryophyta) (Table 1, Supplemental Table 1). These 491 species with expansin homologs are comprised of macroscopic and microscopic organisms widely distributed across the tree of life (Figure 1). In Archaeplastida, expansins are present in red algae (Rhodophyta), which use cellulose as their main cell wall structural carbohydrate, and in *Cyanophora paradoxa*, the sole publicly available Glaucophyta genome sequence (Supplemental Table 2) (Price *et al.*, 2012). Glaucophyta are a rare and largely uncharacterized Archaeplastid group that likely diverged prior to the split between the red (Rhodophyta), and green (Chloroplastida) algal lineages (Price *et al.*, 2012; Adl *et al.*, 2019).

Our search identified few expansin genes in Metazoans (Supplemental Table 3). Tunicates are the only metazoan group known to use cellulose structurally, and acquired their cellulose synthase genes horizontally from bacteria (Nakashima *et al.*, 2004; Sasakura *et al.*, 2005; Sasakura *et al.*, 2016). *Oiklopeura dioica* is the sole tunicate species with a sequenced genome, and it contains an annotated expansin gene. Expansin genes are also annotated in several species of marine bivalves whose diets are partially plant matter or algae. For the sole glaucophyte and the few metazoans with annotated expansin homologs, it remains empirically unconfirmed whether their expansin genes are *bona fide* cellular genes, or sequencing contamination from digestive contents or other plant or microbial DNA (Coutteau & Sorgeloos, 1992; Sakamoto *et al.*, 2007). Many plant pathogenic nematodes have proteins with partial functional and structural overlap with expansins, but their domain structure is reversed compared to the canonical plant and microbial expansin proteins. The evolutionary relationship of nematode expansin-like proteins to plant or microbial expansins remains unclear (Danchin *et al.*, 2010; Cosgrove, 2015; Georgelis *et al.*, 2015), and functional similarities in plant cell wall loosening function may be an example of convergent evolution.

In non-Archaeplastida eukaryotic microbes, Exlx homologs are present in both major lineages of Amoebozoa, one Alveolate (*Vitrella brassicaformis*), one Haptophyte (*Emiliana huxleyi*), and multiple lineages of Stramenopiles (Figure 1, Supplemental Table 2). Few species from these groups have sequenced and well-annotated genomes, and it is likely that more expansin homologs will be identified as more species from these lineages have their genomes sequenced and annotated. *E. huxleyi* and some Stramenopile lineages with expansin genes (such as Phaeophytes) are photosynthetic marine organisms with cellulosic cell walls. Many Amoebozoa and terrestrial Stramenopiles also have cellulosic cell walls. In the slime mold *Dictyostelium discoïdum* (Amoebozoa), expansin genes are expressed while structural cellulose is being rearranged during fruiting body development (Darley *et al.*, 2003). It is likely that *D. discoïdum* – and other microbial eukaryotes with cellulosic cell walls – use expansins to modify their own structural cellulose. Many Oomycetes (a group of non-photosynthetic Stramenopiles) have expansins, but it remains unclear whether oomycetes use expansins for morphogenesis, interactions with plant cell walls, or both. All oomycetes with annotated expansin genes use cellulose structurally and have multiple expansin homologs per genome (Supplemental Table 1) which both suggest functions related to

morphogenesis. However, these same oomycete species colonize plants as hosts and some are among the world's worst agricultural plant pathogens, suggesting possible function(s) related to plant colonization (Helbert *et al.*, 1997; Kamoun, 2001; Hardham, 2007; Grenville-Briggs *et al.*, 2008; Fawke *et al.*, 2015).

Expansin genes were not detected in Archaea in our search, but EXLX homologs are present in an unexpectedly diverse assortment of fungal and bacterial taxa. In fungi, none of which are known to use cellulose structurally, most expansin genes are in Ascomycota, and fewer expansins are in Basidiomycota, Chytrid fungi and symbiotic ectomycorrhizal fungi (Supplemental Table 4). While many bacteria secrete cellulose as part of an extracellular biofilm matrix, none are known to utilize cellulose as a structural component of their cell walls (Schleifer & Kandler, 1972; Branda *et al.*, 2005). Expansins are most abundant in Actinobacteria, Firmicutes, Myxobacteria, γ -Proteobacteria and β -Proteobacteria, and are also present in some Chloroflexi, Bacteroidetes, Cyanobacteria and Verrucomicrobia.

Diversity of ecological niches inhabited by microbes with expansin genes

Since the first discovery of an expansin gene in a bacterium, it was hypothesized that most microbial expansins function as virulence factors by facilitating colonization of plant hosts (Laine *et al.*, 2000; Sampedro & Cosgrove, 2005; Kerff *et al.*, 2008b; Nikolaidis *et al.*, 2013; Cosgrove, 2015). However, we find that only 28% of microbial species with expansins are plant pathogens (Figure 2, Table 2, Supplemental Table 4). More than half (59%, or 290 out of 491) are described as plant commensals, soil inhabitants, or saprophytes – ecological contexts where microbes non-pathogenically interact with live plants and/or decaying plant matter (Lofgren *et al.*, 2018). The remaining 13% of expansin-containing microbes inhabit a variety of terrestrial or aquatic ecological contexts, all of which have in common the presence of live or dead plant or algal matter.

Bacteria comprise 61.3% of microbial species with expansin genes. Especially notable are Myxobacteria, where 85% sequenced species have expansin genes (Table 1), suggesting that their ecological importance as saprophytes has been overshadowed by their use as laboratory models to understand bacterial multicellular behavior (Crespi, 2001; Neil *et al.*, 2005). The Actinobacterial genera *Streptomyces*, *Nocardia*, and *Micromonospora* have many species with expansin genes.

Numerous *Streptomyces* are plant growth promoters or pathogens, but *Nocardia* and *Micromonospora* are predominantly known as soil inhabitants and have few (or no) described plant associations (Supplemental Table 4) (Goodfellow & Williams, 1983; Loria *et al.*, 2006). The high frequency of expansin genes suggests that microbial associations with live plants or dead plant matter is likely more common, and more ecologically important, for these actinobacterial genera than is currently recognized.

Many expansin genes are found in known plant growth promoting rhizobacteria, including strains of *Streptomyces*, *Bacillus*, *Micromonospora*, and *Rhizobacter*. In these species, expansin proteins may function similarly to *B. subtilis* (Kerff *et al.*, 2008b), and increase epiphytic colonization efficiency of plant roots. Only 15% of bacterial species with expansin genes are phytopathogens, and most of those pathogens occur in two γ -proteobacterial lineages, Xanthomonadaceae and Enterobacteriaceae. Other bacterial plant pathogens are sparsely scattered throughout the tree, and include economically important strains of *Ralstonia*, *Acidovorax*, *Streptomyces* and *Clavibacter michiganensis*. A conspicuous number of these expansin-containing bacteria are among the most economically costly agricultural plant pathogens (Table 3). Notably, all expansin-containing bacterial phytopathogens move systemically via xylem at some stage of pathogenesis – an unusual, highly virulent phenotype compared to localized lesions produced by most bacterial plant pathogens (Smith, 1920; Ewald, 1993; Agrios, 2005; Roper, 2011; Mansfield *et al.*, 2012; Shapiro *et al.*, 2012; Bae *et al.*, 2015; Shapiro *et al.*, 2018).

An additional 7% of bacterial species with expansins are marine or freshwater, and likely interact commensally with live algae, or saprophytically degrade dead algal or plant matter. Several species of bacteria with expansin genes were isolated from plants growing in tidal flats, where they may facilitate plant-microbe symbiosis that allow both partners to better tolerate elevated salt levels (Marasco *et al.*, 2012). Expansins were also found from bacteria in acid mine drainage sites, sulfur mats and hot springs (Sgroy *et al.*, 2009; Chung *et al.*, 2015). An expansin gene is present in *Cedecea neteri*, which has been isolated as both a plant commensal and a facultative termite gut symbiont (Farmer *et al.*, 1982; Aguilera *et al.*, 1995; Thong-On *et al.*, 2012; Chan *et al.*, 2014). Expansin genes are found in bacteria (*Paenibacillus*, *Ruminococcus*, *Firmicutes*, *Actinobacteria*) and fungi (*Neocallimastix*, *Anaeromyces*, *Piromyces*, *Aspergillus*) that are commensals in herbivore ruminant

guts and likely aid in degradation of ingested plant matter (Supplemental Table 4) (Morrison *et al.*, 2009).

Only 31.5% of microbes with an expansin homolog are fungi, and they inhabit a more restricted range of ecological habitats than bacteria (Table 2). Almost all fungal species with expansin genes (94%; 146 out of 155) are described as plant pathogens, commensals, or saprophytes, although it is possible this reflects under-sampling of fungi compared to bacteria (Amend *et al.*, 2019). A higher proportion of fungal species with expansins are phytopathogenic (52.3%) compared to the proportion of bacteria that are phytopathogens (15%). Expansin genes are present in many economically devastating fungal pathogens, including many that can cause vascular wilt diseases and can move via xylem during pathogenesis (Figure 2, Table 3, Supplemental Table 4) (Lofgren *et al.*, 2018; Selosse *et al.*, 2018).

Horizontal gene transfer has shaped expansin distribution in microbes

Horizontal transfer of genes between distantly related organisms can introduce new traits and drive rapid evolutionary innovation in the recipient organism. In the microbial expansin phylogeny we identified 21 nodes that are in strong conflict with expected taxonomic relationships (Figure 3), which is suggestive of horizontal gene transfer (HGT). Some of these nodes are well-supported statistically, while others have low support in one or more tests (Table 4), and verifying their placement will require better sampling and/or improved phylogenetic algorithms.

Four nodes represent putative HGT events within Eukaryota (nodes 16, 17, 18, 19), twelve nodes represent putative exchanges within Bacteria (nodes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 21), and five nodes represent putative exchanges between Bacteria and Eukaryota (nodes 12, 13, 14, 15, 20).

Within Eukaryota, the Rhodophyta red alga *Gracilariopsis andersonii* is recovered within Stramenopiles (node 19). Thirteen Ascomycota are recovered in a mixed group with Viridiplantae, the Amoebozoa *Acanthamoeba castellanii* and the Stramenopile *Thalassiosira oceanica* (nodes 16, 17, 18).

Of the 12 within-bacteria HGT events, five involve β -proteobacteria (nodes 1, 5, 7, 11, 21) and four involve γ -proteobacteria (nodes, 1, 2, 3, 8, 11). Two groups of pathogenic γ -proteobacteria – the Xanthomonad and Enterobacterial plant pathogens – group with Firmicutes (nodes 2, 3). Similarities

in ecological habitat and life histories between species at some nodes with putative HGT relationships – such as the marine γ -proteobacteria *Zooshikella ganghwensis* and marine Cyanobacteria at node 8, and plant pathogenic *Ralstonia* within plant pathogenic Xanthomonads at node 1 – suggest ecological niche may be a strong factor driving some expansin HGT events (Smillie *et al.*, 2011). For example, Actinobacterial expansins separate into two main lineages, one comprised mainly of *Streptomyces* and the other mainly of *Micromonospora* and *Nocardia*. These Actinobacterial lineages are separated by a polyphyletic group that includes β -proteobacteria, γ -proteobacteria and Bacteroidetes (nodes 9, 10, 11). The plant commensal *Acidovorax radialis* is part of the β -proteobacteria group recovered within Myxobacteria (node 7), while five other *Acidovorax* that are plant pathogens are in the β -proteobacteria group recovered in Actinobacteria (node 11). All Chloroflexi are recovered within Myxobacteria (nodes 4, 6).

Five of the 21 putative HGT events are exchanges between Bacteria and Eukaryota. Node 15 recovers the expansin gene from *Haloferula* sp. (Verrucomicrobia), a marine symbiont of brown algae (Stramenopiles), near Amoebozoa and Stramenopiles (Mancuso *et al.*, 2016). Node 12 groups Chytrid fungi as sister to Actinobacteria, and places the Chytrid expansins as basal to the bacterial expansins. The Alveolate *Vitrella brassicaformis* (node 14) is recovered as sister to an Actinobacteria-Myxobacteria HGT event (node 13). Node 20 groups the expansin genes from two Actinobacteria (*Hamadaea tsuongensis* and *Streptomyces acidiscabies*) and a β -proteobacteria (*Uliginosibacterium gangwonense*) (node 21) with Stramenopiles. In previous studies, the *Streptomyces acidiscabies* expansin gene was recovered within a group of plant expansins, and this relationship was interpreted as phylogenetic evidence for an HGT from a green land plant donor to a bacterium (Nikolaidis *et al.*, 2013). In our phylogeny – built with a much broader representation of Amoebozoa and Stramenopiles than was available previously – this same *S. acidiscabies* expansin gene, plus an additional expansin gene from an actinobacterium *Hamadaea tsuonensis* – are still an example of Bacteria-Eukaryote HGT, but group with Stramenopiles and not within plants. These Bacteria-Eukaryote HGT events (nodes 12, 13, 14, 15, 20) support the hypothesis that a bacterium could have acquired an expansin gene in a marine environment long before the emergence of land plants ~475-515 million years ago (Morris *et al.*, 2018). None of the 491 microbial expansin genes group within the Viridiplantae, strengthening the hypothesis that land plants were not the expansin gene donors to bacteria and fungi.

Some microbial expansins co-occur with carbohydrate active proteins

In some fungi and bacteria, the two-domain canonical expansin gene is fused to additional glycoside hydrolase (GH) and/or carbohydrate binding module (CBMs). There are currently 83 recognized CBM families. All are non-enzymatic, and often function as part of a larger protein to facilitate adhesion to complex carbohydrates with high substrate specificity (Boraston *et al.*, 2004; Lombard *et al.*, 2013). GHs are a group of enzymes widespread among plants and microbes that degrade complex carbohydrates, and are currently classified into 153 distinct families in the Carbohydrate Active Enzymes database (www.cazy.org). Out of 491 microbial species with expansin genes, 49 (9.9%) exist as fusions to a carbohydrate active domain (Figure 4, Supplemental Table 5). Fifteen of these fusions were previously known (Nikolaidis *et al.*, 2013), and 34 are first identified here.

Carbohydrate binding module family 1 (CBM1) is the only carbohydrate active domain fused to fungal expansins (alternatively referred to in the literature as a ‘swollenins’ (Saloheimo *et al.*, 2002; Brotman *et al.*, 2008; Cosgrove, 2017)). All 14 fungal species with expansin-CBM1 fusions are non-pathogenic. Twelve species of non-pathogenic *Trichoderma*, *Penicillium* and *Talaromyces* form a group distinct from the other predominantly pathogenic fungi without CBM1 fusions, and this group is a well-supported within-Eukaryota HGT event (Figure 3, node 17). Some *Trichoderma* spp., including those with expansin-CBM1 fusions, are among the most thoroughly characterized plant beneficial fungi (Howell, 2003; Harman *et al.*, 2004). We hypothesize that in fungal genetic backgrounds, expansin fusion to CBM1 increases fungal mutualistic capabilities to plant hosts, providing a selective advantage for fungal strains that contain this fusion.

In bacteria, expansins are predominantly found fused to domains from carbohydrate binding module family 2 (CBM2) and/or glycoside hydrolase family 5 (GH5). The Chloroflexi *Herpitosiphon aurantiacus* is the only microbe with an expansin fused to a CBM32 domain. A GH5-expansin fusion construct is present in 15 plant pathogenic Xanthomonadaceae, and previous work indicates multiple independent fusions (Nikolaidis *et al.*, 2013). In some Cyanobacteria, β -proteobacteria, γ -proteobacteria and Actinobacteria, expansins are fused to CBM2 with variable domain arrangements (expansin-CBM2, CBM2-CBM2-expansin or CBM2-expansin; Figure 4). *Clavibacter michiganensis*

is the only species with an expansin domain fused to both GH5 and CBM2 domains (GH5-CBM2-expansin domain arrangement). Most bacteria with unfused expansin genes are not plant pathogens (Figure 2, Supplemental Table 4). However, of the bacterial species with expansin fusions to GH5 and/or CBM2 domains, most (65.3%; 32 out of 49) are virulent phytopathogens (Supplemental Table 5). This suggests that in bacteria, expansin fusions are more likely than unfused expansins to function as a virulence factor (Rocha *et al.*, 2019).

The existence of variable fusion constructs (expansin-CBM32, GH5-expansin, expansin-CBM2, CBM2-expansin, CBM2-CBM2-expansin, and GH5-CBM2-expansin) indicates multiple independent origins of expansin fusions to carbohydrate active domains have occurred in bacteria and fungi. The repeated independent fusions of CBM2 and GH5 domains in bacteria, and only CBM1 in fungi – out of hundreds of CBM and GH families – suggests that CBM1, CBM2 and GH5 active domains in combination with expansin are uniquely useful for bacterial and fungal interactions with cellulosic cell walls.

The enterobacteria may offer mechanistic insight into how fusions can occur. In all enterobacteria, expansin genes are unfused to carbohydrate active domains. However, the plant pathogens *Erwinia tracheiphila* (Shapiro *et al.*, 2016; Shapiro *et al.*, 2018; Rocha *et al.*, 2019) and *Pantoea stewartii* have a canonical expansin gene directly adjacent to – but in a separate open reading frame (ORF) from – a GH5 endoglucanase gene. This expansin-GH5 domain arrangement in *E. tracheiphila* and *P. stewartii* is in opposite positional order to the 15 Xanthomonadaceae with a GH5-expansin fusion construct, suggesting that in either *E. tracheiphila* or *P. stewartii* this gene architecture arose *de novo* and was not acquired horizontally from a Xanthomonadaceae donor. In *E. tracheiphila* and *P. stewartii*, both the expansin and GH5 ORFs have a predicted secretion signal peptide (Rocha *et al.*, 2019). The two coding sequences are separated by a stop codon and a short stretch of 40 nucleotides in *E. tracheiphila* and 51 in *P. stewartii* (KE136322.1, position 16101-17807 in *E. tracheiphila* (Shapiro *et al.*, 2015; Rocha *et al.*, 2019) and NZ_CP017589.1, position 1851-3562 in *P. stewartii*). From this gene architecture, it is possible that either a small deletion in the region between these ORFs, or a mutation in the stop codon separating them, could result in fusion of the two domains into a single gene (Kummerfeld & Teichmann, 2005).

Discussion

We find that microbial expansin genes are more broadly distributed across diverse lineages of bacteria, fungi and other eukaryotic microbes than previously recognized. Especially notable is the presence of expansins in microbes inhabiting a previously unrecognized diversity of terrestrial and aquatic ecological niches, including those not traditionally thought of as cellulose-dominated environments. Many expansin genes are also found in microbes not yet known to interact with plants or algae, suggesting interactions with live or dead plant or algal matter is an overlooked yet important part of their ecological life histories. Identifying expansin genes in such a phylogenetically and ecologically diverse set of microbial species – including many which have not yet been described as interacting with plants or algae – suggests that the immense amount of global cellulose biomass (Bar-On *et al.*, 2018) is an under-recognized selective pressure driving microbial evolution (Smillie *et al.*, 2011).

While the first organism to evolve an expansin gene and the timeframe of this innovation remains unknown, we hypothesize that the original expansin evolved long before the emergence of land plants 475-515 million years ago (Morris *et al.*, 2018). Many microbes that use expansin proteins for cell wall expansion during growth and development – including Stramenopiles, Amoebozoa, Haptophyta, Alveolata, Rhodophyta and Chlorophyta – are lineages much older than land plants. The presence of expansin genes in all eukaryotic organisms with cellulosic cell walls, together with the absence of any extant alternate mechanism for irreversible cellulosic cell wall extension suggests that expansins may have been necessary for the success of the original organism with cellulosic cell walls (Vannerum *et al.*, 2011; Cosgrove, 2015; Van de Poel *et al.*, 2016; Morris *et al.*, 2018). This also raises the possibility that the EXLX microbial expansin subfamily could have been the first to evolve, and then diversified into distinct EXPA, EXPB, EXLA, and EXLB subfamilies in land plants. Ultimately, answering the question of expansin origin and ancient evolutionary dynamics will require greater taxon sampling, high confidence molecular dating of the different lineages that have this gene, and accurate rooting of the expansin phylogeny.

The microbial expansin phylogeny indicates that HGT has been an important process shaping the distribution of expansins among microorganisms, and that expansin gene exchange is ongoing. In some cases, the presence of expansin genes in most sequenced species within a group (such as

Myxobacteria, Xanthomonadaceae, and the *Pectobacterium* and *Dickeya* group of Entobacterial plant pathogens) suggests that original acquisition of an expansin likely occurred in a common ancestor of these taxa before these groups diversified (McDonald & Currie, 2017). In other species – notably several plant pathogens – acquisition of an expansin likely occurred on more recent ecological time scales (Rocha *et al.*, 2019). In several bacterial pathogen species, acquisition of expansin gene or gene fusion resulted in an ability to move systemically via xylem (Rocha *et al.*, 2019) and achieve high within-host titre, which is a high virulence phenotype (Ewald, 1993; Jahr *et al.*, 2000; Mennerat *et al.*, 2010; Pulkkinen *et al.*, 2010; Olarte-Lozano *et al.*, 2014). The high frequency of expansin genes in many virulent fungal and bacterial plant pathogens suggests that expansins or expansin fusions can function as a potent virulence factor when acquired by bacteria and fungi in simplified agro-ecosystems. The increase in virulence conferred by horizontal acquisition of expansins or expansin fusions by microbes in agricultural systems may amount to yet another demonstration of human-driven evolution of pathogenic microorganisms (Mira *et al.*, 2006; Stukenbrock & Bataillon, 2012; McDonald & Stukenbrock, 2016; Shapiro *et al.*, 2018). The amenability of expansin genes to horizontal transfer between phylogenetically divergent microbial lineages, the functionality of this gene in diverse genetic backgrounds, and its repeated occurrence in virulent agricultural pathogens should elicit concern about the possibility of this gene facilitating the emergence of novel, highly virulent pathogen species or strains in managed agricultural settings.

In addition to expansin dissemination via HGT, functional evolution of microbial expansins is likely also driven by fusions with carbohydrate active proteins. There is a correlation between fusion to carbohydrate active domains with a transition between pathogenicity and commensalism, but in opposite directions for bacteria and fungi. In fungi, there was likely only one fusion of an expansin to a CBM1 domain. In bacteria, expansin genes have likely fused multiple times independently with a CBM2 domain, fused at least once to a GH5 domain in Xanthomonadaceae, and appear to be in a possible intermediate arrangement that may result in an additional fusion in *Erwinia tracheiphila* and/or *Pantoea stewartia* (Rocha *et al.*, 2019). The occurrence of expansin fusion constructs across the expansin phylogeny, repeated occurrence of fusions with the same carbohydrate active domains from multiple independent fusion events and the distinct ecological interactions of species with expansin fusions compared to closely related species with unfused expansin genes suggests that these

fusion constructs have emergent (but still unknown) properties beyond their individual constituent domains (Rocha *et al.*, 2019).

We now recognize that all Eukaryotic microbes and macrobes in marine environments – and later in evolutionary history, on land – have evolved as part of complex multi-species communities (Moran *et al.*, 2007; Segev *et al.*, 2016; Chimileski & Kolter, 2017; Kolter & Chimileski, 2019). A mechanism to non-destructively manipulate structural cell wall cellulose would have been highly adaptive for the organism that first evolved a cell wall. This same mechanism could have also been adaptive for the diverse microbes that colonize the surfaces of eukaryotes that have cellulosic cell walls as hosts. The functional flexibility of expansins – which are essential for normal growth and development in some lineages (land plants, red and green algae and some eukaryotic microbes) and accessory in others (bacteria, fungi and possibly other eukaryotic microbes) – appears unique. A more complete understanding of expansin evolutionary origin, functional diversification and emergent properties from fusions with carbohydrate active domains may offer unique insight into the origin of cellulosic cell walls, and mechanisms underlying host-microbe ecological interactions.

Acknowledgements

WRC and DJC were supported by US Department of Energy, Office of Science, Basic Energy Sciences, under grant DE-FG02-84ER13179; LRS was supported by NSF postdoctoral fellowship DBI-1202736; OZ was supported by a Simons Foundation Investigator Award in Mathematical Modeling of Living Systems #327936; JR was supported by Fundación Mexico en Harvard and Conacyt grant 237414. This project used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. We thank Roberto Kolter and the Kolter lab for discussion and insight, and Kerry E. Mauck for critical reading and constructive comments on the manuscript.

Figure Legends

Figure 1. Distribution of expansins within major groups of the Tree of Life. The lineages within Eukaryota and Bacteria that have at least one species with an expansin gene are shown in black, and the lineages without a species with an expansin gene are depicted in gray. Lineages with organisms that use cellulose structurally are marked with a green dot. The phyla relationships are based on (Adl *et al.*, 2012; Pánek *et al.*, 2016; Adl *et al.*, 2019) for Eukaryota and (Hug *et al.*, 2016) for Bacteria.

Figure 2. Ecological niches occupied by microbes with expansins. Each branch of the expansin gene maximum likelihood phylogeny is color-coded according to the ecological life history of that taxon (Supplementary Table 4). The bar charts summarize the distribution of ecological life histories within each major taxonomic group. The tree should be considered unrooted. The scale bar, amino acid substitutions per site. See Supplemental Figure 2 for the tree with the individual taxa labels shown.

Figure 3. Evidence for horizontal exchange of expansin genes within and between Bacteria and Eukaryota. Nodes inferred to be involved in HGT events are shown in bold and numbered 1 through 21, which correspond to their entries in Table 3. Some groups are collapsed to improve presentation; the number of collapsed taxa is shown in parenthesis. Well-supported nodes (Shimodaira–Hasegawa like approximate likelihood ratio test > 70% and/or ultrafast bootstrap > 95%) are marked with black dots. Branches are colored according to taxonomic classification. The tree was reconstructed using maximum likelihood method and should be considered unrooted. Tree scale bar, number of amino acid substitutions per site for the expansin tree.

Figure 4. Gene architecture and phylogenetic distribution of expansin genes that are fused with carbohydrate active domains. In 49 species where the expansin is fused to a carbohydrate active domain, the domain architectures are shown next to the taxa that have them. The tree branches are colored according to taxonomy. The domain architecture diagrams are drawn to scale, with the black line representing the length of the full nucleotide sequences of each gene, and carbohydrate active domains as colored rectangles. The domain architecture scale bar is the length (in nucleotides) of the expansin and carbohydrate active domains. The tree was reconstructed using maximum likelihood

method and should be considered unrooted. The tree scale bar is the number of amino acid substitutions per site.

Supplemental Figure Legends

Supplemental figure 1. Expansin protein structure.

Supplemental figure 2. Phylogenetic relationship among the 600 detected microbial expansins inferred using maximum likelihood method.

Supplemental figure 3. NCBI Common Tree representation of taxonomy for all 491 microbial species with expansin genes.

Supplemental figure 4. Phylogenetic relationships among the 600 detected microbial expansins inferred using Bayesian approach.

Supplemental figure 5. Phylogenetic relationships among the 350 selected microbial expansins inferred using maximum likelihood method.

Supplemental figure 6. Trace plot of two independent runs of Bayesian inference.

Supplemental table 1: List of microbial species with multiple expansin genes per genome.

Supplemental table 2. Number of publicly available genomes in the eukaryotic groups depicted in Figure 1.

Supplemental table 3. Animals whose genomes are known to contain at least one expansin gene.

Supplemental table 4. Ecological metadata for each microbial species with an expansin gene.

Supplemental table 5. Metadata for the 49 expansins that are fused with carbohydrate active domains.

Notes S1. The final microbial expansin gene alignment that contains 689 amino acid sites from 600 unique, *bona fide* non-Viridiplantae expansin proteins from 491 distinct microbial species.

Notes S2. The consensus maximum likelihood tree of the microbial expansin sequences.

Notes S3. Amino acid alignment file of microbial expansin sequences from 350 representative taxa (pruned from the full dataset of 600 microbial sequences)

Notes S4. Tree file (newick format) resulting from the maximum likelihood reconstruction of the 350 pruned taxa.

Notes S5. Full parameters for the Bayesian tree reconstruction.

Notes S6. The majority rule consensus tree file from the sampled trees of MrBayes run 1.

References

- Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F. 2019.** Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* **66**(1): 4-119.
- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V. 2012.** The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* **59**(5): 429-514.
- Agrios. 2005.** *Plant Pathology*. Amsterdam: Elsevier Academic Press.

Aguilera A, Pascual J, Loza E, Lopez J, Garcia G, Liaño F, Quereda C, Ortuño J. 1995.

Bacteraemia with *Cedecea neteri* in a patient with systemic lupus erythematosus.

Postgraduate medical journal **71**(833): 179.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.

Journal of Molecular Biology **215**(3): 403-410.

Amend A, Burgaud G, Cunliffe M, Edgcomb VP, Ettinger CL, Gutiérrez MH, Heitman J, Hom

EFY, Ianiri G, Jones AC, et al. 2019. Fungi in the marine environment: open questions and unsolved problems. *mBIO* **10**(2): e01189-01118.

Bae C, Han SW, Song Y-R, Kim B-Y, Lee H-J, Lee J-M, Yeam I, Heu S, Oh C-S. 2015. Infection

processes of xylem-colonizing pathogenic bacteria: possible explanations for the scarcity of qualitative disease resistance genes against them in crops. *Theoretical and Applied Genetics*

128(7): 1219-1229.

Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on Earth. *Proceedings of the*

National Academy of Sciences **115**(25): 6506-6511.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning

polysaccharide recognition. *Biochemical journal* **382**(3): 769-781.

Branda SS, Vik Å, Friedman L, Kolter R. 2005. Biofilms: the matrix revisited. *Trends in*

Microbiology **13**(1): 20-26.

Brotman Y, Briff E, Viterbo A, Chet I. 2008. Role of swollenin, an expansin-like protein from

Trichoderma, in plant root colonization. *Plant Physiology* **147**(2): 779-789.

Brummell DA, Harpster MH, Civello PM, Palys JM, Bennett AB, Dunsmuir P. 1999.

Modification of expansin protein abundance in tomato fruit alters softening and cell wall polymer metabolism during ripening. *The Plant Cell* **11**(11): 2203-2216.

Burke C, Thomas T, Lewis M, Steinberg P, Kjelleberg S. 2011. Composition, uniqueness and

variability of the epiphytic bacterial community of the green alga *Ulva australis*. *The ISME Journal* **5**(4): 590.

Carey RE, Cosgrove DJ. 2007. Portrait of the expansin superfamily in *Physcomitrella patens*:

comparisons with angiosperm expansins. *Annals of Botany* **99**(6): 1131-1141.

- Carey RE, Hepler NK, Cosgrove DJ. 2013.** *Selaginella moellendorffii* has a reduced and highly conserved expansin superfamily with genes more closely related to angiosperms than to bryophytes. *BMC Plant Biology* **13**(1): 4.
- Carroll G. 1988.** Fungal endophytes in stems and leaves: from latent pathogen to mutualistic symbiont. *Ecology* **69**(1): 2-9.
- Chan K-G, Tan K-H, Yin W-F, Tan J-Y. 2014.** Complete genome sequence of *Cedecea neteri* strain SSMD04, a bacterium isolated from pickled mackerel sashimi. *genomeA* **2**(6): e01339-01314.
- Chen F, Bradford KJ. 2000.** Expression of an expansin is associated with endosperm weakening during tomato seed germination. *Plant Physiology* **124**(3): 1265-1274.
- Chimileski S, Kolter R. 2017.** *Life at the Edge of Sight: A Photographic Exploration of the Microbial World*: Harvard University Press.
- Cho H-T, Cosgrove DJ. 2000.** Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **97**(17): 9783-9788.
- Chung EJ, Park JA, Jeon CO, Chung YR. 2015.** *Gynuella sunshinyii* gen. nov., sp. nov., an antifungal rhizobacterium isolated from a halophyte, *Carex scabrifolia* Steud. *International journal of systematic and evolutionary microbiology* **65**(3): 1038-1043.
- Cosgrove. 2017.** Microbial Expansins. *Annual Review of Microbiology* **71**(1): 479-497.
- Cosgrove D. 2005.** Growth of the plant cell wall. *Nature Reviews* **6**: 850-861.
- Cosgrove D. 2015.** Plant expansins: diversity and interactions with plant cell walls. *Current opinion in plant biology* **25**: 162-172.
- Cosgrove D. 2016.** Catalysts of plant cell wall loosening. *F1000Research* **5**.
- Cosgrove DJ. 2000.** Loosening of plant cell walls by expansins. *Nature* **407**(6802): 321-326.
- Coutteau P, Sorgeloos P. 1992.** The use of algal substitutes and the requirement for live algae in the hatchery and nursery rearing of bivalve molluscs: an international survey. *Journal of Shellfish Research* **II**(2): 467-476.
- Crespi BJ. 2001.** The evolution of social behavior in microorganisms. *Trends in Ecology & Evolution* **16**(4): 178-183.

- Danchin EG, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010.** Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proceedings of the National Academy of Sciences* **107**(41): 17651-17656.
- Darley CP, Li Y, Schaap P, McQueen-Mason SJ. 2003.** Expression of a family of expansin - like proteins during the development of *Dictyostelium discoideum*. *FEBS Letters* **546**(2-3): 416-418.
- Dean R, Van Kan JA, Pretorius ZA, Hammond - Kosack KE, Di Pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J. 2012.** The 'Top 10' fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* **13**(4): 414-430.
- Ding A, Marowa P, Kong Y. 2016.** Genome-wide identification of the expansin gene family in tobacco (*Nicotiana tabacum*). *Molecular genetics and genomics* **291**(5): 1891-1907.
- Egan S, Harder T, Burke C, Steinberg P, Kjelleberg S, Thomas T. 2013.** The seaweed holobiont: understanding seaweed–bacteria interactions. *FEMS Microbiology Reviews* **37**(3): 462-476.
- Ewald PW. 1993.** *Evolution of infectious disease*: Oxford University Press.
- Farmer J, Sheth NK, Hudzinski JA, Rose HD, Asbury MF. 1982.** Bacteremia due to *Cedecea neteri* sp. nov. *Journal of Clinical Microbiology* **16**(4): 775-778.
- Fawke S, Doumane M, Schornack S. 2015.** Oomycete interactions with plants: infection strategies and resistance principles. *Microbiology and molecular biology reviews* **79**(3): 263-280.
- Federhen S. 2011.** The NCBI taxonomy database. *Nucleic Acids Research* **40**(D1): D136-D143.
- Georgelis N, Nikolaidis N, Cosgrove DJ. 2015.** Bacterial expansins and related proteins from the world of microbes. *Applied Microbiology and Biotechnology* **99**(9): 3807-3823.
- Georgelis N, Yennawar NH, Cosgrove DJ. 2012.** (PDB: 4FER) Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proceedings of the National Academy of Sciences*: 201213200.
- Goodfellow M, Williams S. 1983.** Ecology of Actinomycetes. *Annual Reviews in Microbiology* **37**(1): 189-216.
- Gray-Mitsumune M, Mellerowicz EJ, Abe H, Schrader J, Winzél A, Sterky F, Blomqvist K, McQueen-Mason S, Teeri TT, Sundberg B. 2004.** Expansins abundant in secondary xylem belong to subgroup A of the α -expansin gene family. *Plant Physiology* **135**(3): 1552-1564.

- Grenville-Briggs LJ, Anderson VL, Fugelstad J, Avrova AO, Bouzenzana J, Williams A, Wawra S, Whisson SC, Birch PR, Bulone V. 2008.** Cellulose synthesis in *Phytophthora infestans* is required for normal appressorium formation and successful infection of potato. *The Plant Cell* **20**(3): 720-738.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**(3): 307-321.
- Guy L, Kultima JR, Andersson SG. 2010.** genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**(18): 2334-2335.
- Haney CH, Samuel BS, Bush J, Ausubel FM. 2015.** Associations with rhizosphere bacteria can confer an adaptive advantage to plants. *Nature Plants* **1**(6).
- Hardham AR. 2007.** Cell biology of plant–oomycete interactions. *Cellular Microbiology* **9**(1): 31-39.
- Harman GE, Howell CR, Viterbo A, Chet I, Lorito M. 2004.** *Trichoderma* species—opportunistic, avirulent plant symbionts. *Nature Reviews Microbiology* **2**(1): 43.
- Helbert W, Sugiyama J, Ishihara M, Yamanaka S. 1997.** Characterization of native crystalline cellulose in the cell walls of Oomycota. *Journal of Biotechnology* **57**(1-3): 29-37.
- Hématy K, Cherk C, Somerville S. 2009.** Host–pathogen warfare at the plant cell wall. *Current opinion in plant biology* **12**(4): 406-413.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2017.** UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular biology and evolution*: msx281.
- Howell C. 2003.** Mechanisms employed by *Trichoderma* species in the biological control of plant diseases: the history and evolution of current concepts. *Plant Disease* **87**(1): 4-10.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hermsdorf AW, Amano Y, Ise K. 2016.** A new view of the tree of life. *Nature Microbiology* **1**: 16048.
- Hwang IS, Oh E-J, Lee HB, Oh C-S. 2019.** Functional characterization of two cellulase genes in the Gram-positive pathogenic bacterium *Clavibacter michiganensis* for wilting in tomato. *Molecular Plant-Microbe Interactions* **32**(4): 491-501.

- Im K-H, Cosgrove DJ, Jones AM. 2000.** Subcellular localization of expansin mRNA in xylem cells. *Plant Physiology* **123**(2): 463-470.
- Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R. 2000.** The endo- β -1, 4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. *Molecular Plant-Microbe Interactions* **13**(7): 703-714.
- Junior AT, Dolce LG, de Oliveira Neto M, Polikarpov I. 2015.** Xanthomonas campestris expansin-like X domain is a structurally disordered beta-sheet macromolecule capable of synergistically enhancing enzymatic efficiency of cellulose hydrolysis. *Biotechnology letters* **37**(12): 2419-2426.
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermin LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**(6): 587.
- Kamoun S. 2001.** Nonhost resistance to *Phytophthora*: novel prospects for a classical problem. *Current opinion in plant biology* **4**(4): 295-300.
- Katoh K, Standley DM. 2013.** MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**(4): 772-780.
- Kende H, Bradford K, Brummell D, Cho H-T, Cosgrove D, Fleming A, Gehring C, Lee Y, McQueen-Mason S, Rose J. 2004.** Nomenclature for members of the expansin superfamily of genes and proteins. *Plant molecular biology* **55**(3): 311-314.
- Kerff F, Amoroso A, Herman R, Sauvage E, Petrella S, Filée P, Charlier P, Joris B, Tabuchi A, Nikolaidis N. 2008a.** PDB ID: 3D30 Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proceedings of the National Academy of Sciences* **105**(44): 16876-16881.
- Kerff F, Amoroso A, Herman R, Sauvage E, Petrella S, Filée P, Charlier P, Joris B, Tabuchi A, Nikolaidis N, et al. 2008b.** Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proceedings of the National Academy of Sciences* **105**(44): 16876-16881.
- Kimura S, Itoh T. 1995.** Evidence for the role of the glomerulocyte in cellulose synthesis in the tunicate, *Metandrocarpa uedai*. *Protoplasma* **186**(1-2): 24-33.

- Kolter R, Chimileski S. 2019.** The end of microbiology. *Environmental Microbiology* **20**(6): 1955-1959.
- Kummerfeld SK, Teichmann SA. 2005.** Relative rates of gene fusion and fission in multi-domain proteins. *Trends in Genetics* **21**(1): 25-30.
- Laine MJ, Haapalainen M, Wahlroos T, Kankare K, Nissinen R, Kassuwi S, Metzler MC. 2000.** The cellulase encoded by the native plasmid of *Clavibacter michiganensis* ssp. *sepedonicus* plays a role in virulence and contains an expansin-like domain. *Physiological and Molecular Plant Pathology* **57**(5): 221-233.
- Lee D-K, Ahn JH, Song S-K, Do Choi Y, Lee JS. 2003.** Expression of an expansin gene is correlated with root elongation in soybean. *Plant Physiology* **131**(3): 985-997.
- Lee Y, Choi D, Kende H. 2001.** Expansins: ever-expanding numbers and functions. *Current opinion in plant biology* **4**(6): 527-532.
- Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ. 2002.** Plant expansins are a complex multigene family with an ancient evolutionary origin. *Plant Physiology* **128**(3): 854-864.
- Li Y, Jones L, McQueen-Mason S. 2003.** Expansins and cell growth. *Current opinion in plant biology* **6**(6): 603-610.
- Lofgren LA, LeBlanc NR, Certano AK, Nachtigall J, LaBine KM, Riddle J, Broz K, Dong Y, Bethan B, Kafer CW. 2018.** *Fusarium graminearum*: pathogen or endophyte of North American grasses? *New Phytologist* **217**(3): 1203-1212.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2013.** The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**(D1): D490-D495.
- Loria R, Kers J, Joshi M. 2006.** Evolution of plant pathogenicity in *Streptomyces*. *Annual review of phytopathology* **44**: 469-487.
- Mancuso FP, D'Hondt S, Willems A, Airoidi L, De Clerck O. 2016.** Diversity and temporal dynamics of the epiphytic bacterial communities associated with the canopy-forming seaweed *Cystoseira compressa* (Esper) Gerloff and Nizamuddin. *Frontiers in Microbiology* **7**(476).

- Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA. 2012.** Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* **13**(6): 614-629.
- Marasco R, Rolli E, Ettoumi B, Vigani G, Mapelli F, Borin S, Abou-Hadid AF, El-Behairy UA, Sorlini C, Cherif A. 2012.** A drought resistance-promoting microbiome is selected by root system under desert farming. *PloS one* **7**(10): e48479.
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR. 2016.** CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research* **45**(D1): D200-D203.
- Marchler-Bauer A, Bryant SH. 2004.** CD-Search: protein domain annotations on the fly. *Nucleic Acids Research* **32**(suppl_2): W327-W331.
- McDonald BA, Stukenbrock EH. 2016.** Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philosophical transactions - Royal Society. Biological sciences* **371**(1709): 20160026.
- McDonald BR, Currie CR. 2017.** Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *mBIO* **8**(3): e00644-00617.
- Meibom KL, Li XB, Nielsen AT, Wu CY, Roseman S, Schoolnik GK. 2004.** The *Vibrio cholerae* chitin utilization program. *Proceedings of the National Academy of Sciences* **101**(8): 2524.
- Mennerat A, Nilsen F, Ebert D, Skorping A. 2010.** Intensive farming: evolutionary implications for parasites and pathogens. *Evolutionary Biology* **37**(2-3): 59-67.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013.** Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**(5): 1188-1195.
- Mira A, Pushker R, Rodríguez-Valera F. 2006.** The Neolithic revolution of bacterial genomes. *Trends in Microbiology* **14**(5): 200-206.
- Moran M, Belas R, Schell M, Gonzalez J, Sun F, Sun S, Binder B, Edmonds J, Ye W, Orcutt B. 2007.** Ecological genomics of marine Roseobacters. *Applied and Environmental Microbiology* **73**(14): 4559-4569.

- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018.** The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences* **15**(10): E2274-E2283.
- Morrison M, Pope PB, Denman SE, McSweeney CS. 2009.** Plant biomass degradation by gut microbiomes: more of the same or something new? *Current Opinion in Biotechnology* **20**(3): 358-363.
- Nakashima K, Yamada L, Satou Y, Azuma J-i, Satoh N. 2004.** The evolutionary origin of animal cellulose synthase. *Development Genes and Evolution* **214**(2): 81-88.
- Neil RB, Hite D, Kelrick MI, Lockhart ML, Lee K. 2005.** Myxobacterial biodiversity in an established oak-hickory forest and a savanna restoration site. *Current Microbiology* **50**(2): 88-95.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**(1): 268-274.
- Nikolaidis N, Doran N, Cosgrove DJ. 2013.** Plant expansins in bacteria and fungi: evolution by horizontal gene transfer and independent domain fusion. *Molecular Biology and Evolution* **31**(2): 376-386.
- Niu B, Paulson JN, Zheng X, Kolter R. 2017.** Simplified and representative bacterial community of maize roots. *Proceedings of the National Academy of Sciences* **114**(12): E2450-E2459.
- Ogasawara S, Shimada N, Kawata T. 2009.** Role of an expansin - like molecule in *Dictyostelium morphogenesis* and regulation of its gene expression by the signal transducer and activator of transcription protein Dd - STATa. *Development, growth & differentiation* **51**(2): 109-122.
- Olarte-Lozano M, Mendoza-Nuñez MA, Pastor N, Segovia L, Folch-Mallol J, Martínez-Anaya C. 2014.** PcEx11 a novel acid expansin-like protein from the plant pathogen *Pectobacterium carotovorum*, binds cell walls differently to BsEXLX1. *PLoS ONE* **9**(4): e95638.
- Pánek T, Zadrobníková E, Walker G, Brown MW, Gentekaki E, Hroudová M, Kang S, Roger AJ, Tice AK, Vlček Č. 2016.** First multigene analysis of Archamoebae (Amoebozoa: Conosa) robustly reveals its phylogeny and shows that Entamoebidae represents a deep lineage of the group. *Molecular Phylogenetics and Evolution* **98**: 41-51.

- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004.** UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**(13): 1605-1612.
- Pien S, Wyrzykowska J, McQueen-Mason S, Smart C, Fleming A. 2001.** Local expression of expansin induces the entire process of leaf development and modifies leaf shape. *Proceedings of the National Academy of Sciences* **98**(20): 11812-11817.
- Popper ZA, Michel G, Hervé C, Domozych DS, Willats WG, Tuohy MG, Kloareg B, Stengel DB. 2011.** Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual Review of Plant Biology* **62**: 567-590.
- Popper ZA, Tuohy MG. 2010.** Beyond the green: understanding the evolutionary puzzle of plant and algal cell walls. *Plant Physiology* **153**(2): 373-383.
- Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber AP, Schwacke R, Gross J, Blouin NA, Lane C. 2012.** *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* **335**(6070): 843-847.
- Pulkkinen K, Suomalainen L-R, Read A, Ebert D, Rintamäki P, Valtonen E. 2010.** Intensive fish farming and the evolution of pathogen virulence: the case of columnaris disease in Finland. *Proceedings of the Royal Society of London B: Biological Sciences* **277**(1681): 593-600.
- R. 2015.** R: A language and environment for statistical computing. *R Foundation of Statistical Computing Vienna, Austria*(<https://www.R-project.org>).
- Ramakrishna P, Duarte PR, Rance GA, Schubert M, Vordermaier V, Dai Vu L, Murphy E, Barro AV, Swarup K, Moirangthem K. 2019.** EXPANSIN A1-mediated radial swelling of pericycle cells positions anticlinal cell divisions during lateral root initiation. *Proceedings of the National Academy of Sciences* **116**(17): 8597-8602.
- Rocha J, Shapiro LR, Kolter R. 2019.** A horizontally acquired expansin gene increases virulence of the emerging plant pathogen *Erwinia tracheiphila*. *bioRxiv*: 681643.
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* **61**(3): 539-542.

- Roper MC. 2011.** *Pantoea stewartii* subsp. *stewartii*: lessons learned from a xylem - dwelling pathogen of sweet corn. *Molecular Plant Pathology* **12**(7): 628-637.
- Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, et al. 2017.** The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Research* **45**: D271-D281.
- Sakamoto K, Touhata K, Yamashita M, Kasai A, Toyohara H. 2007.** Cellulose digestion by common Japanese freshwater clam *Corbicula japonica*. *Fisheries Science* **73**(3): 675-683.
- Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyssönen E, Bhatia A, Ward M, Penttilä M. 2002.** Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European Journal of Biochemistry* **269**(17): 4202-4211.
- Sampedro J, Cosgrove DJ. 2005.** The expansin superfamily. *Genome Biology* **6**(12): 242.
- Santiago TR, Pereira VM, de Souza WR, Steindorff AS, Cunha BA, Gaspar M, Fávares LC, Formighieri EF, Kobayashi AK, Molinari HB. 2018.** Genome-wide identification, characterization and expression profile analysis of expansins gene family in sugarcane (*Saccharum* spp.). *PLoS ONE* **13**(1): e0191081.
- Sasakura Y, Nakashima K, Awazu S, Matsuoka T, Nakayama A, Azuma J-i, Satoh N. 2005.** Transposon-mediated insertional mutagenesis revealed the functions of animal cellulose synthase in the ascidian *Ciona intestinalis*. *Proceedings of the National Academy of Sciences* **102**(42): 15134-15139.
- Sasakura Y, Ogura Y, Treen N, Yokomori R, Park S-J, Nakai K, Saiga H, Sakuma T, Yamamoto T, Fujiwara S. 2016.** Transcriptional regulation of a horizontally transferred gene from bacterium to chordate. *Proceedings of the Royal Society B: Biological Sciences* **283**(1845): 20161712.
- Schipper O, Schaefer D, Reski R, Fleming A. 2002.** Expansins in the bryophyte *Physcomitrella patens*. *Plant molecular biology* **50**(4-5): 789-802.
- Schleifer KH, Kandler O. 1972.** Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological reviews* **36**(4): 407.

- Segev E, Wyche TP, Kim KH, Petersen J, Ellebrandt C, Vlamakis H, Barteneva N, Paulson JN, Chai L, Clardy J, et al. 2016. Dynamic metabolic exchange governs a marine algal-bacterial interaction. *eLife* 5: e17473.
- Selosse MA, Schneider - Maunoury L, Martos F. 2018. Time to re - think fungal ecology? Fungal ecological niches are often prejudged. *New Phytologist* 217(3): 968-972.
- Sgroy V, Cassán F, Masciarelli O, Del Papa MF, Lagares A, Luna V. 2009. Isolation and characterization of endophytic plant growth-promoting (PGPB) or stress homeostasis-regulating (PSHB) bacteria associated to the halophyte *Prosopis strombulifera*. *Applied Microbiology and Biotechnology* 85(2): 371-381.
- Shabat SKB, Sasson G, Doron-Faigenboim A, Durman T, Yaacoby S, Miller MEB, White BA, Shterzer N, Mizrahi I. 2016. Specific microbiome-dependent mechanisms underlie the energy harvest efficiency of ruminants. *The ISME Journal* 10(12): 2958.
- Shapiro L, De Moraes CM, Stephenson AG, Mescher MC. 2012. Pathogen effects on vegetative and floral odours mediate vector attraction and host exposure in a complex pathosystem. *Ecology Letters* 15(12): 1430-1438.
- Shapiro LR, Paulson JN, Arnold BJ, Scully ED, Zhaxybayeva O, Pierce N, Rocha J, Klepac-Ceraj V, Holton K, Kolter R. 2018. An introduced crop plant is driving diversification of the virulent bacterial pathogen *Erwinia tracheiphila*. *mBio* 9(5): e01307-01318.
- Shapiro LR, Scully ED, Roberts D, Straub TJ, Geib SM, Park J, Stephenson AG, Rojas ES, Liu Q, Beattie G, et al. 2015. Draft genome sequence of *Erwinia tracheiphila*, an economically important bacterial pathogen of cucurbits. *genomeA* 3(3): e00482-00415.
- Shapiro LR, Scully ED, Straub TJ, Park J, Stephenson AG, Beattie GA, Gleason ML, Kolter R, Coelho MC, Moraes CMD, et al. 2016. Horizontal gene acquisitions, mobile element proliferation, and genome decay in the host-restricted plant pathogen *Erwinia tracheiphila*. *Genome Biology and Evolution* 8(3): 649-664.
- Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480(7376): 241-244.

- Smith EF. 1920.** *An introduction to bacterial diseases of plants*. Philadelphia: W.B. Saunders Company.
- Stukenbrock EH, Bataillon T. 2012.** A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. *PLoS Pathogens* **8**(9): e1002893.
- Tamai N, Tatsumi D, Matsumoto T. 2004.** Rheological properties and molecular structure of tunicate cellulose in LiCl/1, 3-dimethyl-2-imidazolidinone. *Biomacromolecules* **5**(2): 422-432.
- Tancos MA, Lowe - Power TM, Peritore - Galve FC, Tran TM, Allen C, Smart CD. 2018.** Plant - like bacterial expansins play contrasting roles in two tomato vascular pathogens. *Molecular Plant Pathology* **19**(5): 1210-1221.
- Thong-On A, Suzuki K, Noda S, Inoue J-i, Kajiwara S, Ohkuma M. 2012.** Isolation and characterization of anaerobic bacteria for symbiotic recycling of uric acid nitrogen in the gut of various termites. *Microbes and environments* **27**(2): 186-192.
- Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, et al. 2014.** XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **16**(5): 62-74.
- Van de Poel B, Cooper ED, Van Der Straeten D, Chang C, Delwiche CF. 2016.** Transcriptome profiling of the green alga *Spirogyra pratensis* (Charophyta) suggests an ancestral role for ethylene in cell wall metabolism, photosynthesis, and abiotic stress responses. *Plant Physiology* **172**(1): 533-545.
- Vannerum K, Huysman MJ, De Rycke R, Vuylsteke M, Leliaert F, Pollier J, Lütz-Meindl U, Gillard J, De Veylder L, Goossens A. 2011.** Transcriptional analysis of cell growth and morphogenesis in the unicellular green alga *Micrasterias* (Streptophyta), with emphasis on the role of expansin. *BMC Plant Biology* **11**(1): 128.
- Wang T, Park YB, Caporini MA, Rosay M, Zhong L, Cosgrove DJ, Hong M. 2013.** Sensitivity-enhanced solid-state NMR detection of expansin's target in plant cell walls. *Proceedings of the National Academy of Sciences* **110**(41): 16444-16449.
- Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ. 2006a.** Crystal structure and activities of EXPB1 (*Zea m 1*), a β -expansin and group-1 pollen allergen from maize. *Proceedings of the National Academy of Sciences* **103**(40): 14664-14671.

Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ. 2006b. (PDB: 2HCZ) Crystal structure and activities of EXPB1 (*Zea m 1*), a β -expansin and group-1 pollen allergen from maize. *Proceedings of the National Academy of Sciences* **103**(40): 14664-14671.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**(W1): W445-W451.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**(1): 28-36.

Table 1: Distribution of microbial expansins in major taxonomic groups

Group	Number of named species	Number of species with a genome sequence	Total number of expansin genes
Myxobacteria (Deltaproteobacteria)	66	26	22
Xanthomonads (Gammaproteobacteria)	220	86	19
Actinobacteria	3244	1058	174
Enterobacteria (Gammaproteobacteria)	312	207	14
Betaproteobacteria	753	364	16
Firmicutes	2717	1475	46
Amoebozoa	792	40	35
Stramenopiles	3233	67	54
Basidiomycetes	1776	250	14
Ascomycetes	17565	683	174
Other taxa	Undetermined	Undetermined	33

Table 2: Summary of the raw counts and proportion of bacteria and fungi (out of a total of 456 bacterial and fungal species) that occupy different ecological niches

	Fungi	Bacteria	Combined Fungi and Bacteria
Plant Pathogen	76 (49%)	46 (15%)	138 (28%)
Plant Commensal	11 (7%)	42 (13.95%)	53 (10.7%)
Soil	19 (12.3%)	139 (46.2%)	158 (32.2%)
Saprophyte	43 (27.7%)	36 (11.96%)	79 (16.1%)
Freshwater	0	4 (1.3%)	4 (0.8%)
Gut Microbe	5 (3.2%)	5 (1.7%)	10 (2%)
Hot Spring or Sulfur Mat	0	5 (1.7%)	5 (1.0%)
Marine	0	22 (7.3%)	22 (4.5%)
Wastewater	0	2 (0.6%)	2 (0.4%)
Total	155 (31.5%)	301 (61.3%)	456

Table 3: Presence of expansin genes in the ‘Top 10’ most important species of plant pathogenic bacteria and fungi (Dean *et al.*, 2012; Mansfield *et al.*, 2012).

Top 10 most important bacterial plant pathogens	Microbial expansin gene	Accession Number	Top 10 most important fungal plant pathogens	Microbial expansin gene	Accession Number
<i>Pseudomonas syringae</i>	No	—	<i>Magnaporthe oryzae</i>	Yes	XP_003711472
<i>Ralstonia solanacearum</i>	Yes	AOE88952	<i>Botrytis cinerea</i>	Yes	CCD54783
<i>Agrobacterium tumefaciens</i>	No	—	<i>Fusarium graminearum</i>	Yes	CEF78987
<i>Xanthomonas oryzae</i>	Yes	WP_011409499	<i>Fusarium oxysporum</i>	Yes	KNB11838
<i>Xanthomonas campestris</i>	Yes	WP_011038627	<i>Blumeria graminis</i>		—
<i>Xanthomonas axonopodis</i>	Yes	WP_048490938	<i>Mycosphaerella graminicola</i> [†]	Yes	EGP90221
<i>Erwinia amylovora</i>	No	—	<i>Ustilago maydis</i>		—
<i>Xylella fastidiosa</i>	Yes	SHG20409	<i>Melampsora lini</i> [‡]		—
<i>Dickeya (dadantii and solani)</i>	Yes	WP_038924663, KHN53802	<i>Puccinia spp.</i>	Yes	KNZ59084, EFP85064, KNF04021
<i>Pectobacterium (carotovorum and atrosepticum)</i>	Yes	SHG09776, AIK14148	<i>Colletotrichum spp.</i>	Yes	OLN88490, CCF37207, KXH45263, XP_007594649, KXH43193, KXH26982, ENH88520, EQB51143,

					KDN66674
--	--	--	--	--	----------

Table 4: Maximum likelihood and Bayesian support values testing statistical strength at 21 nodes representing putative horizontal gene transfer (HGT) events

Node Number	HGT recipient	HGT recipient groups		Full ML tree support (★SH-aLRT/◆UFboot)	Full Bayesian tree support (● Posterior Probability)	Pruned ML tree support (★SH-aLRT/◆UFboot)
		within or is sister-grouped to	Intra- or Inter-domain HGT			
1	β-proteobacteria (<i>Ralstonia</i> spp.)	γ-proteobacteria (Xanthomonadaceae)	Bacteria-Bacteria	97.9/100	100	98/100
2	γ-proteobacteria (Xanthomonadaceae)	Firmicutes	Bacteria-Bacteria	59.2/84	73	39.2/40
3	γ-proteobacteria (Enterobacteria)	Firmicutes	Bacteria-Bacteria	44.3/85	Polytomy	21.9/84
4	Chloroflexi (<i>Roseiflexus castenholzii</i> , <i>Oscillochloris trichoides</i> , <i>Chloroflexus islandicus</i> , <i>Chloroflexus aurantiacus</i>)	Δ-Proteobacteria	Bacteria-Bacteria	52.4/92	Polytomy	29.9/59
5	β-proteobacteria (<i>Sphaerotilus natans</i>)	Δ-Proteobacteria	Bacteria-Bacteria	89.6/99	Polytomy	93.3/93

6	Δ -Proteobacteria (<i>Plesiocystis pacifica</i>)	Chloroflexi (<i>Herpetosiphon aurantiacus</i> , <i>Herpetosiphon geysericola</i>)	Bacteria-Bacteria	97.5/99	99	96.9/96
7	β -proteobacteria (<i>Janthinobacterium sp.</i> , <i>Methylibium sp.</i> , <i>Acidovorax radialis</i> , <i>Leptothrix cholodnii</i> , <i>Polyangium brachysporum</i>)	Δ -Proteobacteria	Bacteria-Bacteria	93.5/99	97	92.2/97
8	γ -proteobacteria (<i>Zooshikella ganghwensis</i>)	Cyanobacteria	Bacteria-Bacteria	64.6/100	98	9.9/89
9	Polyphyletic group of β -proteobacteria, Bacteroidetes, γ - proteobacteria	Actinobacteria	Bacteria-Bacteria	85.1/90	Polytomy	84.9/91

10	Bacteroidetes (<i>Aquimarina agarivorans</i> , <i>Aquimarina agarilytica</i> , <i>Aquimarina latercul</i>)	β -proteobacteria and γ -proteobacteria	Bacteria-Bacteria	97.4/100	99	97.1/98
11	β -proteobacteria (several <i>Acidovorax</i> spp., <i>Methylibium</i> , <i>Rhizobacter</i> , <i>Mitsuaria</i>)	γ -proteobacteria	Bacteria-Bacteria	97.8/100	96	97.6/100
12	Chytrid anaerobic gut fungi (<i>Neocallimastix</i> <i>californiae</i> , <i>Anaeromyces robustus</i> , <i>Piromyces finnim</i> , <i>Allomyces macrogyneu</i>)	Actinobacteria	Eukaryota- Bacteria	87.2/98	76	93.7/89
13	Alveolate (<i>Vitrella</i> <i>brassicaformis</i>), Δ - Proteobacteria (<i>Sorangium cellulosum</i>)	Actinobacteria	Eukaryota- Bacteria	87.4/99	52	83.8/89
14	Alveolata (<i>Vitrella</i> <i>brassicaformis</i>)	Δ -Proteobacteria (<i>Sorangium</i> <i>cellulosum</i>)	Eukaryota- Bacteria	81.8/98	N/A[§]	N/A[§]

15	Verrucomicrobia (<i>Haloferula sp.</i>)	Amoebozoa, Stramenopiles	Eukaryota- Bacteria	96/98	100	95.9/100
16	Fungal expansin-CBMI fusion proteins (Ascomycetes)	Viridiplantae	Eukaryota- Eukaryota	97.2/98	100	95.1/99
17	Amoebozoa (<i>Acanthamoeba castellanii</i>), Stramenopiles (<i>Thalassiosira oceanica</i>)	Fungal expansin-CBMI fusion proteins (Ascomycetes)	Eukaryota- Eukaryota	44.9/98	Polytomy	36.2/95
18	Amoebozoa (<i>Acanthamoeba castellanii</i>)	Stramenopiles (<i>Thalassiosira oceanica</i>)	Eukaryota- Eukaryota	82/99	N/A[§]	N/A[§]
19	Rhodophyta (<i>Gracilariopsis andersonii</i>)	Stramenopiles	Eukaryota- Eukaryota	85.5/98	92	99.9/100
20	Actinobacteria (<i>Hamadaea tsuonensis</i> , <i>Streptomyces acidiscabie</i>) β- proteobacteria (<i>Uliginosibacterium</i>)	Amoebozoa	Eukaryota- Bacteria	91.5/93	79	84.7/86

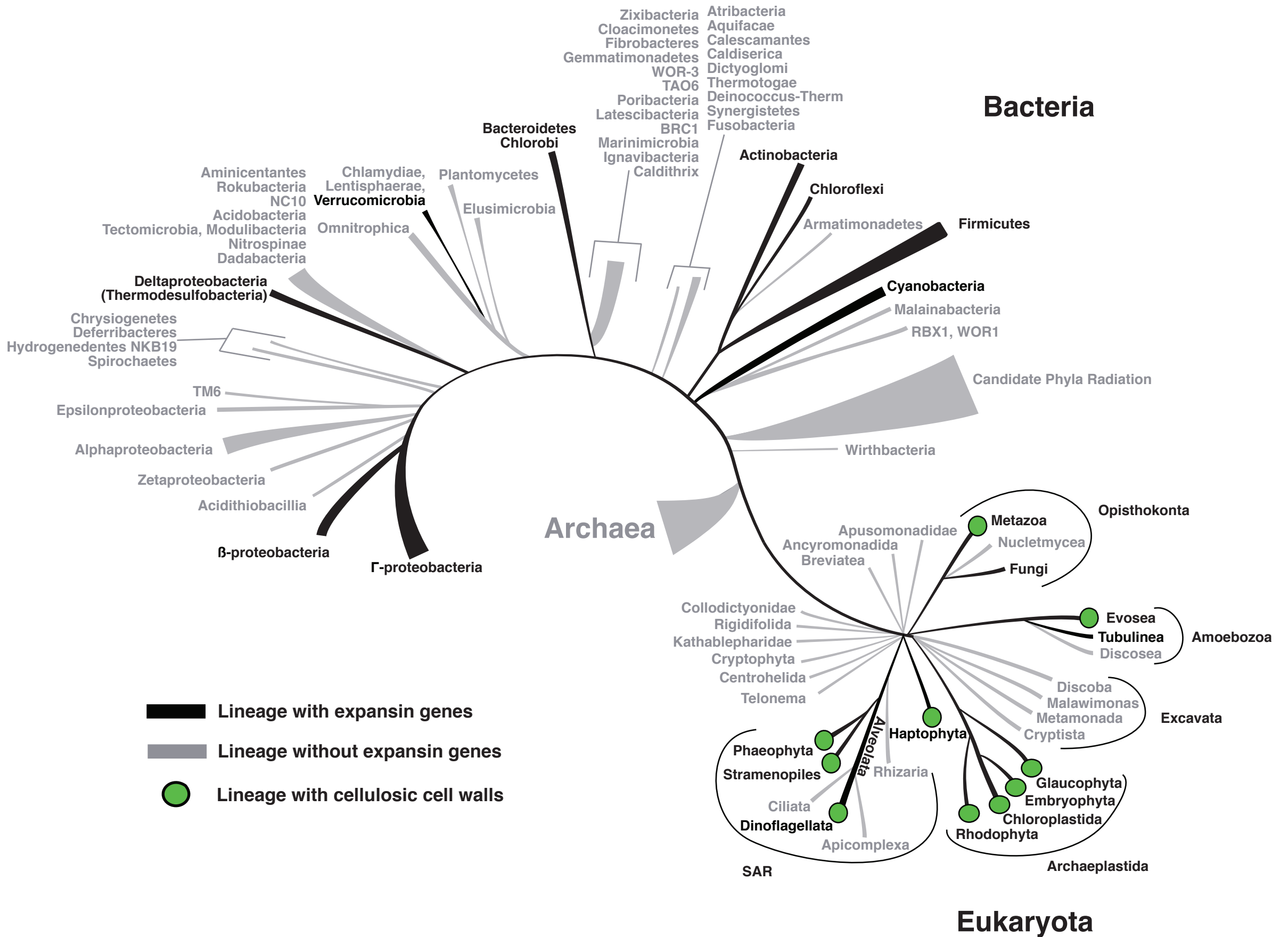
	<i>gangwonense</i>)					
21	β -proteobacteria (<i>Uliginosibacterium</i> <i>gangwonense</i>)	Actinobacteria (<i>Hamadaea</i> <i>tsuonensis</i> , <i>Streptomyces</i> <i>acidiscabies</i>)	Bacteria-Bacteria	79.6/98	92	N/A[§]

★ SH-aLRT, SH-like approximate likelihood ratio test (as in Guindon 2010). SH-aLRT significance level is over 75

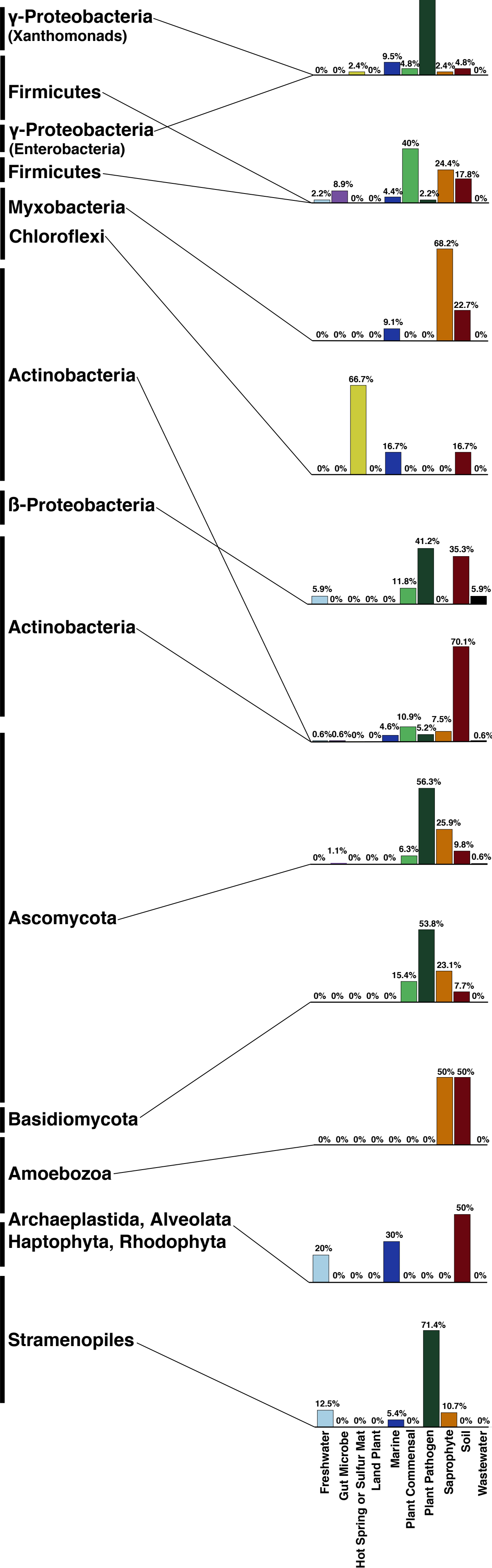
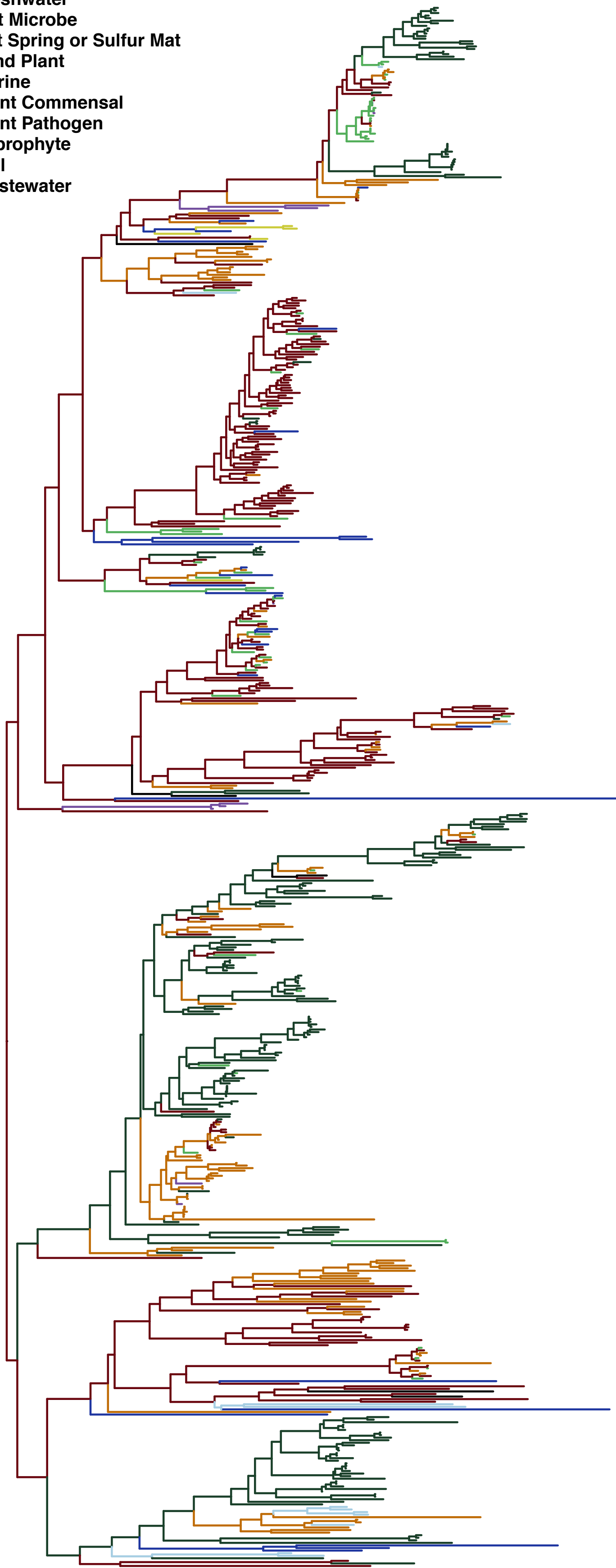
◆ UFboot, ultra-fast bootstrap approximation approach (as in Minh 2013). Significance level is over 95

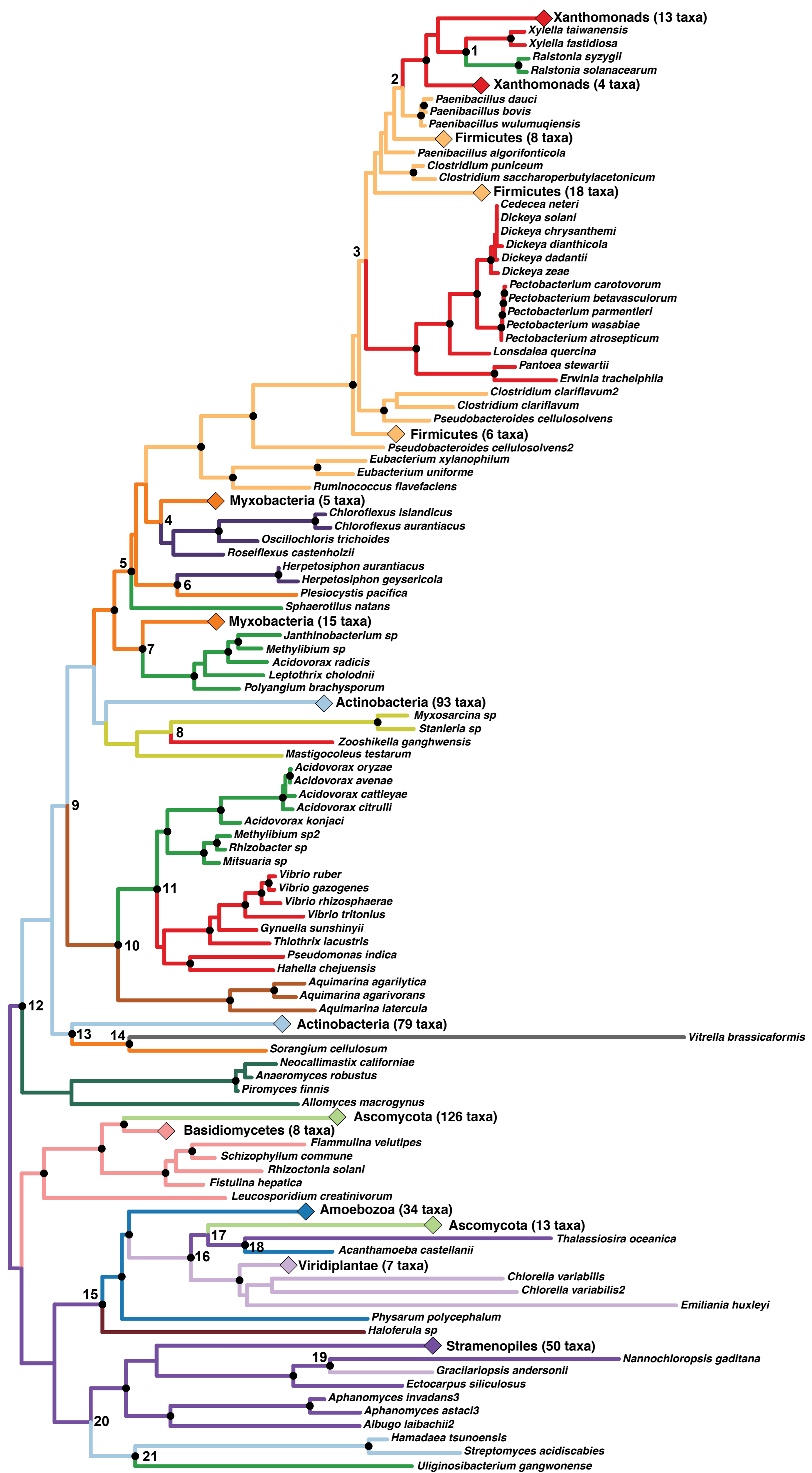
● Bayesian posterior probability. Considered well-supported over 95

§ Taxa on long branches that were removed from the pruned ML tree



- Ecology**
- Freshwater
 - Gut Microbe
 - Hot Spring or Sulfur Mat
 - Land Plant
 - Marine
 - Plant Commensal
 - Plant Pathogen
 - Saprophyte
 - Soil
 - Wastewater





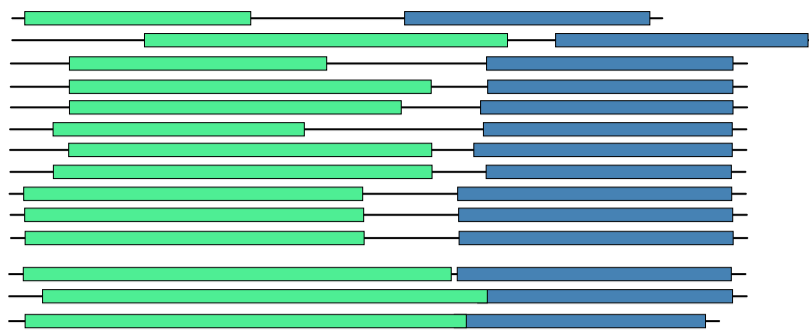
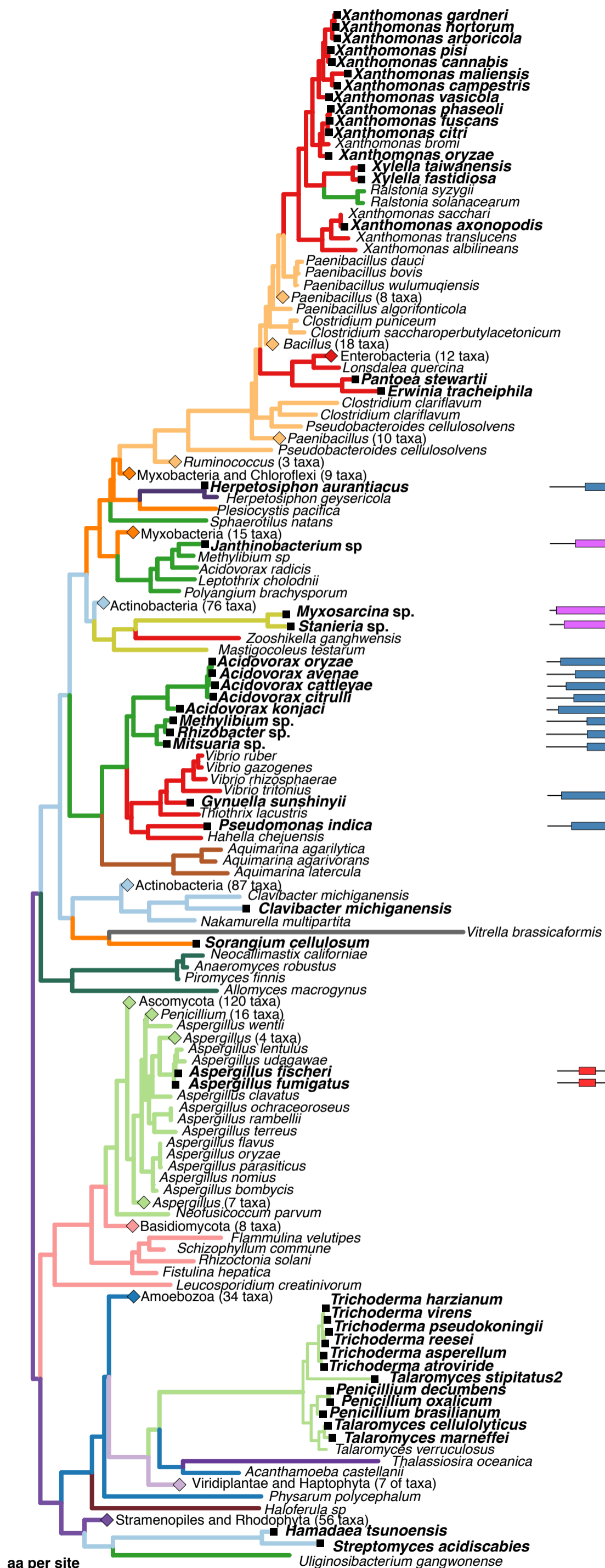
- Taxonomic Group**
- γ-Proteobacteria
 - β-Proteobacteria
 - Firmicutes
 - Myxococcales
 - Cyanobacteria
 - Bacteroidetes
 - Chloroflexi
 - Actinobacteria
 - Alveolata
 - Chytrid fungi
 - Ascomycetes
 - Basidiomycetes
 - Amoebozoa
 - Stramenopiles
 - Verrucomicrobia
 - Viridiplantae, Rhodophyta and Haptophyta

Bacteria

Mixed Bacteria/Eukaryotes

Eukaryotes

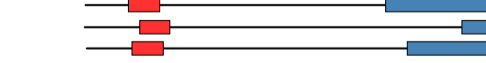
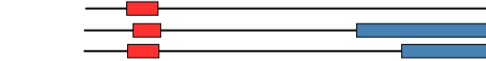
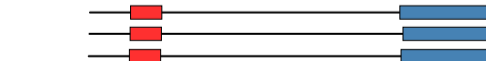
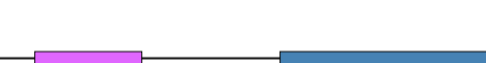
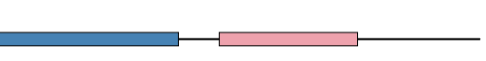
Mixed Bacteria/Eukaryotes



Domain

- █ GH5
- █ Expansin
- █ CBM32
- █ CBM2
- █ CBM1

100 nt
Fusion construct scale



Taxonomic Group

- █ β-Proteobacteria
- █ Firmicutes
- █ γ-Proteobacteria
- █ Myxobacteria
- █ Chloroflexi
- █ Actinobacteria
- █ Cyanobacteria
- █ Bacteroidetes
- █ Alveolata
- █ Chytrid fungi
- █ Ascomycetes
- █ Basidiomycetes
- █ Amoebozoa
- █ Stramenopiles
- █ Archaeplastida and Haptophyta
- █ Verrucomicrobia