PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

Duplication and neofunctionalization of a horizontally transferred xyloglucanase as a facet of the Red Queen coevolutionary dynamic

Victoria Attah[a], David S. Milner[a], Yufeng Fang[b,1], Xia Yan[c], Guy Leonard[a], Joseph Heitman[b], Nicholas J. Talbot[c], and Thomas A. Richards[a,2]

Edited by David Baulcombe, University of Cambridge, Cambridge, United Kingdom; received November 13, 2022; accepted April 20, 2024

Oomycete protists share phenotypic similarities with fungi, including the ability to cause plant diseases, but branch in a distant region of the tree of life. It has been suggested that multiple horizontal gene transfers (HGTs) from fungi-to-oomycetes contributed to the evolution of plant-pathogenic traits. These HGTs are predicted to include secreted proteins that degrade plant cell walls, a barrier to pathogen invasion and a rich source of carbohydrates. Using a combination of phylogenomics and functional assays, we investigate the diversification of a horizontally transferred xyloglucanase gene family in the model oomycete species *Phytophthora sojae*. Our analyses detect 11 xyloglucanase paralogs retained in *P. sojae*. Using heterologous expression in yeast, we show consistent evidence that eight of these paralogs have xyloglucanase function, including variants with distinct protein characteristics, such as a long-disordered C-terminal extension that can increase xyloglucanase activity. The functional variants analyzed subtend a phylogenetic node close to the fungi-to-oomycete transfer, suggesting the horizontally transferred gene was a bona fide xyloglucanase. Expression of three xyloglucanase paralogs in *Nicotiana benthamiana* triggers high-reactive oxygen species (ROS) generation, while others inhibit ROS responses to bacterial immunogens, demonstrating that the paralogs differentially stimulate pattern-triggered immunity. Mass spectrometry of detectable enzymatic products demonstrates that some paralogs catalyze the production of variant breakdown profiles, suggesting that secretion of variant xyloglucanases increases efficiency of xyloglucan breakdown as well as diversifying the damage-associated molecular patterns released. We suggest that this pattern of neofunctionalization and the variant host responses represent an aspect of the Red Queen host–pathogen coevolutionary dynamic.

horizontal gene transfer | gene duplication | host parasite interaction | plant cell wall

Significance

The oomycetes are a diverse group of eukaryotic microbes that include devastating pathogens of plants. Oomycetes perceive, invade, and colonize plants in similar ways to fungi, in part because they acquired genes from fungi. Many of these genes have undergone and retained multiple rounds of gene duplication. One key enzyme for attacking plant cell wall structures is called xyloglucanase. Oomycete xyloglucanase has undergone multiple gene duplications, leading to variants including enzymes with protein extensions that increase enzymatic activity. Some xyloglucanase variants trigger unique patterns of reactive oxygen species (ROS) in planta, and generate different profiles of cell wall breakdown products—such outcomes could act to overwhelm the plant immune system, enabling oomycete pathogens to proliferate.

Oomycetes are heterotrophic protists that form part of the stramenopile lineage—also sometimes called heterokonts (1–5). They resemble fungi in their filamentous growth and osmotrophic feeding (6), which initially led to their taxonomic placement within the kingdom Fungi (7). It is now known that oomycetes and fungi branch on distant parts of the eukaryotic tree of life (8–13), suggesting that biological similarities of the two groups were the product of convergent evolution (14, 15). The oomycetes are a highly diverse group of protists, but include ecologically and agriculturally destructive pathogens of plants (16). Many plant pathogens penetrate or break down the cell walls of their hosts, a heterogeneous refractory structure rich in diverse polysaccharides. Plant cell walls are composed of cellulose cross-linked by glycans and embedded within a matrix of pectin, hemicelluloses, lignin, other aromatic polymers, and proteins, which together provide structural integrity (17, 18). One of the most abundant hemicellulosic polysaccharides is xyloglucan, consisting of a backbone of D-glucose linked by β-1-4 glycosidic bonds (similar to cellulose); however, most are substituted with α-1,6-linked xylose (often in turn capped with galactose, arabinose, or fucose). Owing to the different possible variations of the sugar side chains on xylosyl residues, the xyloglucan structure varies significantly between plant species (19–23). A nomenclature-based system has been used to define side chain variants and describe the oligosaccharides released by enzymatic degradation of xyloglucan (24). Interestingly, mixed xyloglucan oligosaccharides (obtained through enzymatic extraction from apple pomace) have also been demonstrated to act as damage-associated molecular patterns (DAMPs) in *Vitis vinifera* (grapevine) and *Arabidopsis thaliana*, leading to direct activation of a range of plant immune responses across phylogenetically distant plant species, which aim to halt or limit pathogen colonization (25).

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: Department of Plant Health, GreenLight Biosciences Inc, Durham, NC 27709.

[2]To whom correspondence may be addressed. Email: thomas.richards@biology.ox.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218927121/-/DCSupplemental.

Published June 3, 2024.

The genomes of many plant pathogens, especially hemibiotrophic species, are replete with genes encoding secreted glycoside hydrolases, predicted to break down external plant cell wall layers (26, 27). These gene families include previously identified horizontal gene transfers (HGTs) from fungi-to-oomycetes and include xyloglucanase (28–34). HGT, or lateral gene transfer (LGT), describes the movement of genes across species boundaries, a process which can potentially add new functionality to the recipient lineage (see refs. 35 and 36). New functionality can also be acquired through gene duplication events—specifically when a gene duplication is coupled with neofunctionalization of sister paralogs. In oomycetes, there is evidence of widespread gene duplication (post-acquisition) of fungal-derived HGT genes (31, 37), resulting in multiple paralogs, many of which are predicted to encode secreted products which interact with the host. This pattern of evolution seems to be prevalent among economically important *Phytophthora* species, including *Phytophthora sojae*—a hemibiotrophic pathogen of soybean (31, 38, 39).

We know very little regarding the functional significance of HGT events and subsequent gene family duplications in *Phytophthora* species. Gene duplication events can be linked to subsequent adaptation regarding increased gene dosage and/or functional divergence known as neofunctionalization (40) or subfunctionalization (41, 42). Previous work has demonstrated that one paralog of the horizontally acquired *P. sojae* Glycoside Hydrolase 12 (GH12) gene family, PsXLP1, has undergone a significant pattern of deletion within the encoded open reading frame, rendering a catalytically inactivated protein that acts as a "decoy" of plant immune response, helping to facilitate infection (43). The PsXLP1 gene is located on the genome in a head-to-head orientation with the closely related paralogous PsXEG1 gene and has been shown to inhibit the host immune protein GmGIP1, whereas none of the other 10 GH12 members tested were shown to inhibit the same host protein (43). Additional work has focused on detailed examination of the functional PsXEG1 paralog and its contribution to *P. sojae* virulence (44). These results are consistent with neofunctionalization through peptide sequence evolution being adaptive with regard to *P. sojae* virulence and immune system evasion.

Here, we characterize all 11 HGT-derived paralogs that have resulted from gene duplication in the HGT-derived *P. sojae* GH12 gene family, using bioinformatics, heterologous protein expression in *Saccharomyces cerevisiae*, and transient expression in *Nicotiana benthamiana*. Our analysis suggests a pattern of functional divergence linked to unique protein characteristics important for xyloglucanase function. We show that some paralogs produce different xyloglucan breakdown profiles and suggest that secretion of multiple variants could therefore increase the efficacy of cell wall substrate breakdown. Importantly, we demonstrate that variant paralogs trigger distinct patterns of reactive oxygen species (ROS) accumulation in planta, suggesting that the host must be on high alert to multiple related protein signatures during pathogen interaction. The concordant action of multiple paralogs transcribed during infection may therefore act to perturb plant immune perception both within individual infections (for example, via differential transcription), and over evolutionary divergence (by replacement of variant paralog functions, for instance). Such a scenario would suggest that the pattern of gene loss, duplication, and neofunctionalization observed here is an evolutionary consequence of the Red Queen coevolutionary dynamic (45, 46), which predicts that host–pathogen coevolution acts to increase genetic variation by generating diversity of molecular signatures involved in pathogenic interactions (47).

Results

The aim of this work was to investigate the functional consequences of a series of gene duplications of a horizontally transferred, putatively secreted xyloglucanase, previously identified as a candidate HGT from fungi to the oomycetes primarily based on taxon distribution data (29, 31, 33). In this study, we tested the hypothesis that gene duplication events, which have occurred after the HGT event, drove diversification, and neofunctionalization of this enzyme function in *P. sojae*. To better understand the dynamics of GH12 gene turnover in oomycetes, we first explored patterns of lineage-specific gene gains (e.g., via duplication and fixation) or losses (e.g., via inactivation and deletion) across representative *Phytophthora* spp., and *Hyaloperonospora arabidopsidis* (an obligate biotrophic pathogen of *A. thaliana*) taxa (Fig. 1*A*). A total of 11 paralogs of GH12 in *P. sojae* were confirmed by HMM (hiddenMarkov model) searches (48–50). The *P. sojae* GH12 paralogs were aligned to a range of candidate GH12 genes identified in additional oomycete genomes sampled (Fig. 1*A*) using the same HMM search method described above but excluding some sequences representing incomplete gene models (*Materials and Methods*). We then constructed a phylogenetic tree to explore the specific evolutionary relationships between *P. sojae* GH12 paralogs, using focused sampling of orthologs in closely related oomycetes, demonstrating the branching of the 11 *P. sojae* genes across multiple paralog clades (Fig. 1*A*). We note that this analysis shows weak phylogenetic resolution in several sections of the tree (i.e., low bootstrap support values) and clear evidence of multiple cases of gene loss both within the nine resolved clades identified (Fig. 1*A*), and unspecified gene losses among the intermediate phylogenetic groups which lack resolution. Even with many nodes lacking strong support, these data demonstrate that the gene family has been subject to numerous cases of gene duplication and loss typical of strong diversifying selection operating on the GH12 gene repertoire and putatively contributing to functional variation.

FungiDB (55, 56) was used to locate approximate genomic coordinates of each *P. sojae* GH12 gene—the relative approximate distances between the paralogs are shown in kilobases (kb) (Fig. 1*E*). These results demonstrate two phylogenetically defined sister paralog pairs, which also demonstrate close clustering on chromosomal contigs (*P. sojae*_559651 (PsXEG1) and *P. sojae*_360375 (PsXLP1) on PHYSO_scaffold_4; *P. sojae*_247788 and *P. sojae*_482953 on PHYSO_scaffold_2). Close chromosomal positions can be used as additional evidence to support sister-paralog relationships identified in phylogenetic trees. The genetic mechanisms that led to these numerous GH12 gene duplications in *P. sojae* are currently unclear.

We searched *P. sojae* life-cycle transcriptome datasets [FungiDB (55, 56)] to identify and compare how the eleven *P. sojae* GH12 paralogs were expressed across mycelial, cyst, and 3 d postinfection (soybean hypocotyls infected with *P. sojae* strain P6497). Analysis of these transcriptomes demonstrated that the duplicated sequences exhibit stage-specific expression, as well as differences in levels of expression [fragments per kilobase of exon model per million mapped reads (FPKM) were used to calculate Log2 values (Fig. 1*B*)]. Four paralogs are expressed in *P. sojae* mycelium, seven paralogs are expressed in cysts, and nine paralogs are expressed during plant infection. Four of the eleven xyloglucanases are expressed to some level during all three stages: *P. sojae*_360375 (PsXLP1), *P. sojae*_559651 (PsXEG1), *P. sojae*_482953, and *P. sojae*_355355 (Fig. 1*B*). Gene expression of two of the eleven xyloglucanases was undetectable in all stages: *P. sojae*_247788 and *P. sojae*_520599 (Fig. 1*B*). We note that previous analyses describe

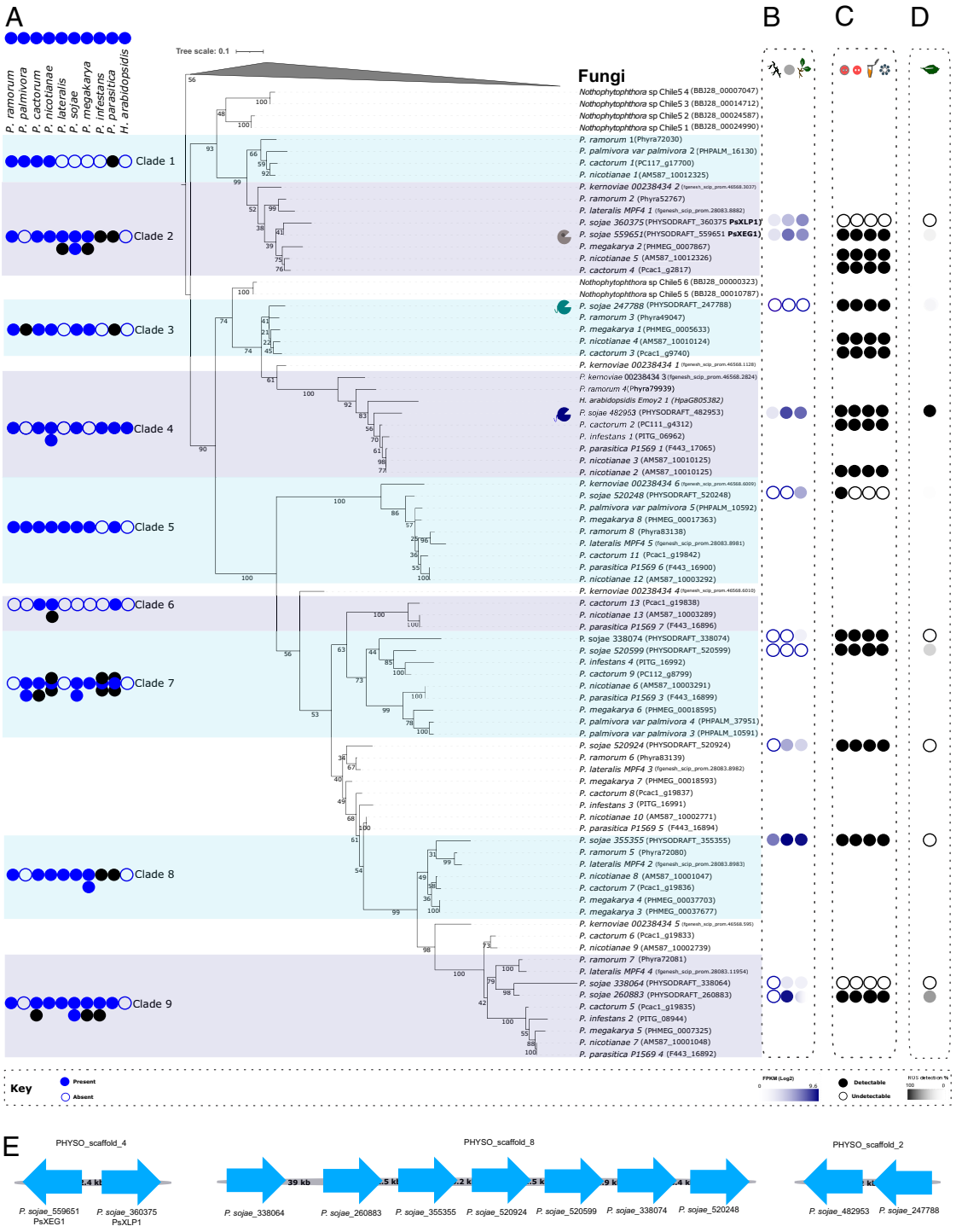

Fig. 1. Evolutionary history of oomycete GH12. (*A*) Maximum likelihood tree constructed with IQ-Tree v2.0.3 [WAG+R5 model of evolution (51) selected as the best-fit model by ModelFinder (52)]. The tree was constructed from an alignment of 161 sequences comprising 211 amino acids to confirm the evolutionary relationships between oomycete GH12. The final tree was visualized with iTOL (53), and nodes indicate results (%) of nonparametric bootstrap (200 pseudoreplicates) (54). The tree is rooted with a fungal outgroup because fungi were previously identified as the putative donor group of the GH12 HGT into the oomycetes (31, 33). The eleven *P. sojae* GH12 paralogs are visually identifiable as they have adjacent transcriptome profiles (column B). Symbols indicate the *P. sojae* paralog with a structurally inferred additional binding site (gray), and paralogs with C-terminal tails 1 and 2 (blue and green, respectively). The numbers of GH12 orthologs across representative *Phytophthora* spp., and *H. arabidopsidis* (an obligate pathogen of *A. thaliana*) genomes are shown to the *Left* of the phylogenetic tree, indicating lineage-specific putative local duplications or losses. Some are colored black, indicating they are absent in the tree due to exclusion because of gappy gene models, or they formed uninformative branches, potentially affecting bootstrap resolution (i.e., very long branches or very short indicating little protein variation). All *P. sojae* paralogs were retained. (*B*) *P. sojae* life-cycle transcriptome data [FungiDB: (55, 56)] were used to identify how the eleven *P. sojae* GH12 paralogs were expressed across (column 1) mycelial, (column 2), cyst and (column 3) 3 d postinfection (soybean hypocotyls infected with *P. sojae* strain P6497)—expressed in FPKM(Log2). *C. Xyloglucanase* functional data generated during this study from (column 1) cell culture agar plate-based enzyme assay, (column 2) concentrated supernatant agar-plate based enzyme assay, (column 3) concentrated supernatant DNS reducing sugar assay, and (column 4) concentrated supernatant mass spectrometry of xyloglucan breakdown products (black fill = function detected, white fill = no function detected). (*D*) ROS data in response to *P. sojae* xyloglucanase variants, presented as percentage ROS generation (%) of the variant that triggered the highest ROS accumulation in *N. benthamiana* (i.e., *P. sojae*_482953, which is presented as 100%). (*E*) FungiDB (55, 56) was used to locate approximate genomic coordinates of each *P. sojae* GH12 gene—the relative approximate distances between the paralogs are shown in kilobases (kb), providing additional support for the sister paralog relationships in the phylogenetic tree for: *P. sojae*_559651 PsXEG1 and *P. sojae*_360375 PsXLP1 (PHYSO_scaffold_4); and *P. sojae*_482953 (PHYSO_scaffold_2). Scaffold numbers are shown in the figure.

fine-scale correlated expression patterns for *P. sojae*_360375 (PsXLP1) and *P. sojae*_559651 (PsXEG1), consistent with their expression during all stages described here (43, 44). The authors demonstrate that both variants are expressed at low levels in mycelium, cyst, germinating cyst, and zoospore stages, with an increase in expression at 10 min postinfection of soybean hypocotyls, up to the highest peak in expression at 1 h postinfection, before declining (43). We also note that publicly available *P. sojae* transcriptome datasets are limited in capturing just one timepoint during infection, potentially missing important expression patterns during this phase—however, we include the data for completeness.

Amino acid alignments demonstrated the predicted amino acid sequences of *P. sojae*_482953 and *P. sojae*_247788 possess an extended C terminus compared to the other nine paralogs, while *P. sojae*_559651 (PsXEG1) has two single amino acid insertions not found in other *P. sojae* GH12 paralogs (*SI Appendix*, Fig. S1). This analysis also confirmed that *P. sojae*_360375 (PsXLP1) has a curtailed protein sequence in comparison to the other eleven *P. sojae* paralogs, as previously documented (43).

Heterologous Expression of *P. sojae* GH12 Paralogs in *S. cerevisiae* Demonstrates that Nine Enzymes Exhibit Activity against Xyloglucan. All *P. sojae* GH12 gene models were manually checked using publicly available RNA-seq data for *P. sojae* (55, 56). Confirmed candidate nucleotide sequences for all 11 *P. sojae* GH12 paralogs were synthesized, codon-optimized for *S. cerevisiae*, and cloned into plasmid p426-GPD for expression in *S. cerevisiae*. To maximize heterologous protein secretion in yeast, *P. sojae* N-terminal signal peptide sequences were replaced with the native *S. cerevisiae* mating factor α (MFA) signal peptide sequence (57) (*SI Appendix*, Table S1). Concentrated supernatants of yeast strains, each expressing a *P. sojae* GH12 paralog, were used as crude protein extracts for enzyme activity assays using 3,5-dinitrosalicylic acid (DNS) reagent compared to a yeast strain carrying an empty vector. DNS is an aromatic compound that is reduced in the presence of reducing sugars (released during the breakdown of polysaccharides), to 3-amino-5-nitrosalicylic acid, which absorbs light at ~540 nm (58). Therefore, an increase in absorbance at 544 nm over time was used to infer enzymatic activity.

After 6 h of incubation with 1% (w/v) xyloglucan (pH 7, 30 °C), significant levels of reducing sugars released by enzyme activity were detected for the *P. sojae*_482953, 559651 (PsXEG1), 520924, and 338074 xyloglucanase paralogs (Dunnett's test; Fig. 2*A*). After prolonged incubation (72 h), significant levels of reducing sugars were also detected for the *P. sojae*_260883, 355355, 520599, and 247788 enzymes (Dunnett's test; Fig. 2*B*), indicating that eight out of eleven paralogs display enzymatic activity toward xyloglucan compared to a control yeast carrying an empty vector. However, these data demonstrate that the paralogs have very different rates of catalysis, consistent with a pattern of diversified function. Interestingly, *P. sojae*_482953 and *P. sojae*_559651 (PsXEG1) show more rapid degradation than the other active paralogs (Fig. 2*A*). As both paralogs subtend the ancestral node, at the point when HGT from the fungi to the oomycetes was predicted to occur, they therefore putatively share the conserved function of the ancestral fungal-to-oomycete HGT, i.e., the HGT was likely a bona fide xyloglucanase. The shared derived function of this gene family is further suggested by previous studies which report homologous fungal proteins that function as xyloglucanases (59). These results confirm that detectable xyloglucanase function was, however, lost during expansion of this gene family in the branch leading to *P. sojae*_360375 (PsXLP1). Loss of enzymatic function in *P. sojae*_360375 (PsXLP1) has previously been demonstrated to be linked to a deletion event within the peptide-encoding sequence and shown to result in improved host immune-evasion, demonstrating a link between evolution of peptide characteristics and an evolutionary response to host function (43). We also find loss of enzymatic function in the branch leading to *P. sojae*_338064, but note that this paralog expressed in *S. cerevisiae* was subject to manual sequence correction, due to identification of three miscalled introns that were removed from the predicted gene sequence (*SI Appendix*, Fig. S8). It is therefore possible that this paralog does not accurately reflect the mature transcript expressed by *P. sojae* in vivo, or indeed that this variant is a possible pseudogene. This result is consistent

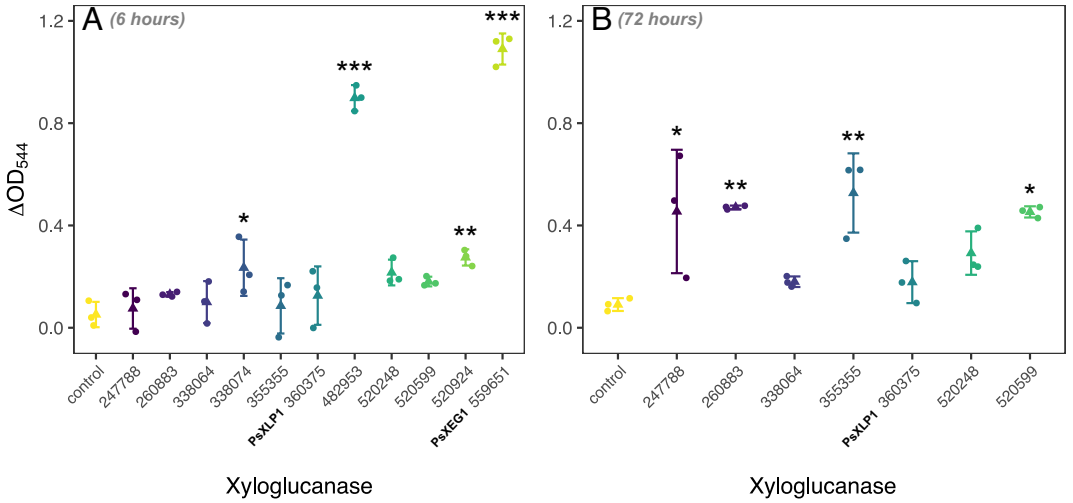

Fig. 2. Comparative assay of xyloglucanase function across the 11 *P. sojae* GH12 paralogs. *P. sojae* GH12 paralogs secreted into *S. cerevisiae* culture supernatants were incubated with 1% (w/v) xyloglucan at 30 °C, pH 7; an increase in absorbance (OD_{544}) of DNS reagent added to the samples is suggestive of an increase in the reducing sugars released (i.e., from the breakdown of the substrate). (*A*) Significant enzymatic activity toward xyloglucan was detected up to 6 h of incubation for *P. sojae*_338074, 482953, 520924, and 559651 xyloglucanase variants (Dunnett's test). (*B*) After 72 h of incubation, significant levels of reducing sugars were also detected for the *P. sojae*_247788, 260883, 355355, and 520599 enzymes (Dunnett's test), indicating that eight out of the eleven paralogs display enzymatic activity toward xyloglucan. *P. sojae*_482953 and *P. sojae*_559651 appeared to show more rapid degradation than the other active paralogs under these conditions. *P. sojae*_338074, 482953, 520924, and 559651 enzymatic activities toward xyloglucan were not shown in figure B due to their strong function at 6 h. No significant reducing sugars were detected in the vector-only sample (n = 3, ±SD).

with the absence of enzymatic function for this protein. We note that *P. sojae*_520248 displayed enzymatic activity to xyloglucan in a cell culture agar plate-based assay, confirming function for a ninth paralog (Fig. 1*C* and *SI Appendix*, Fig. S6*B*), but it was not possible to detect function of this paralog from concentrated supernatants, potentially falling below our limit of detection for enzymatic activity.

Both *P. sojae*_482953 and *P. sojae*_559651 (PsXEG1), although representing functional forms that subtend the ancestral HGT phylogenetic node, showed radically different protein sequence modifications (i.e., in the case of *P. sojae*_559651 (PsXEG1), two separate single amino acid insertions and, in the case of *P. sojae*_482953, a C-terminal extension (*SI Appendix*, Fig. S1). These results make it difficult to identify how peptide characteristics relate to xyloglucanase function or the ancestral function of the horizontally transferred encoded gene, but confirm that this gene family has been subject to additional peptide level variation beyond the deletion identified by Ma et al. (2017) for *P. sojae*_360375 (PsXLP1) (43). To explore this puzzle further, we conducted a series of additional assays to investigate how variation in these peptide characteristics relates to function.

***P. sojae*_482953 and the Orthologs in *Phytophthora cactorum* and *Phytophthora nicotianae* Have a Highly Disordered C-terminal Extension which Influences Catalysis.** *P. sojae*_482953 and the orthologs identified from *P. cactorum* and *P. nicotianae* encode disordered C-terminal extensions (*SI Appendix*, Fig. S1). As such, the protein structures were unable to be accurately modeled using Phyre2 (60) or AlphaFold (61, 62). To investigate the functional significance of the extensions at the C terminus of *P. sojae*_482953 and orthologs from *P. cactorum* and *P. nicotianae*, truncated versions of the proteins (removing the C-terminal regions) were engineered and expressed in *S. cerevisiae*. Following incubation with 1% (w/v) xyloglucan (pH 7, 30 °C) for 6 h, we find that removal of the extension significantly impairs xyloglucanase function for *P. sojae*_482953 (*t* test; *P*-value = 0.04) and its ortholog in *P. cactorum* (*t* test; *P*-value = 0.0001), but this reduction in enzymatic activity was not found to be significant for the *P. nicotianae* ortholog (*t* test; *P*-value > 0.05) at any timepoint (Fig. 3*A*).

***P. sojae*_247788 and the Orthologs in *P. cactorum* and *P. nicotianae* Also Have a Highly Disordered C-terminal Extension.** *P. sojae*_247788 is the most closely related GH12 paralog to *P. sojae*_482953, forming a sister relationship with 74% bootstrap support (Fig. 1*A*), and lying in close proximity to this gene on the chromosomal contig (Fig. 1*E*). Interestingly, like *P. sojae*_482953, *P. sojae*_247788 (and orthologs in *P. cactorum* and *P. nicotianae*) also encode a long, disordered C-terminal extension. Similarly, these amino acid extensions were unable to be accurately modeled as a protein structure using Phyre2 (60) or AlphaFold (61, 62). Furthermore, it was not possible to identify any regions of consistently alignable sequence homology between the *P. sojae*_482953 and *P. sojae*_247788 C-terminal tail extensions. Interestingly, we detected low enzymatic activity of *P. sojae*_247788 toward xyloglucan, suggesting reduced function in this branch (Fig. 2). Prolonged incubation of *P. sojae*_247788 (full-length and truncated) (72 h), demonstrated no significant differences between these two variant protein forms (*SI Appendix*, Fig. S7). To further explore function of this paralog, we repeated our experimental assay for orthologous proteins of *P. sojae*_247788 sampled from *P. cactorum* and *P. nicotianae*. We note that the C-terminal tail extensions were also present in the *P. cactorum* and *P. nicotianae* orthologous proteins, but they showed considerable amino acid

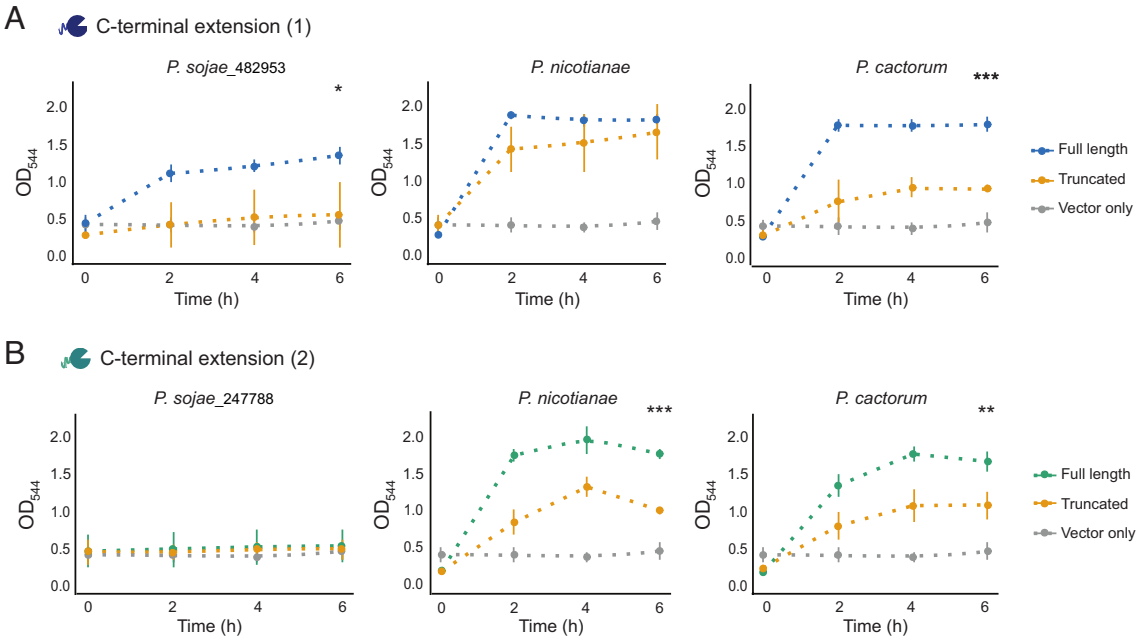

Fig. 3. Exploration of function among xyloglucanase paralogs with C-terminal extensions. (*A*) *P. sojae*_482953 (full-length) and *P. sojae*_482953 (truncated), and orthologs in *P. cactorum* and *P. nicotianae* secreted into *S. cerevisiae* culture supernatants were incubated with 1% (w/v) xyloglucan at 30 °C, pH 7; an increase in absorbance (OD_{544}) of DNS reagent added to the samples is suggestive of an increase in the reducing sugars released (i.e., from the breakdown of the substrate). Following incubation with xyloglucan for 6 h, we find that removal of the C-terminal extension significantly impairs xyloglucanase function for *P. sojae*_482953 (*t* test; *P*-value = 0.04) and its ortholog in *P. cactorum* (*t* test; *P*-value = 0.0001), but this reduction was not found to be significant for the *P. nicotianae* ortholog at any timepoint. No significant reducing sugars were detected in the vector-only sample (n = 3, ±SD). (*B*) *P. sojae*_247788 (truncated) and *P. sojae*_247788 (full-length), and orthologs in *P. cactorum* and *P. nicotianae* secreted into *S. cerevisiae* culture supernatants were incubated with 1% (w/v) xyloglucan at 30 °C, pH 7; an increase in absorbance (OD_{544}) of DNS reagent added to the samples is suggestive of an increase in the reducing sugars released (i.e., from the breakdown of the substrate). *P. sojae*_247788 (full-length) gave weak enzymatic activity toward xyloglucan by this method, but the orthologous proteins of *P. cactorum* and *P. nicotianae* demonstrated significantly higher xyloglucanase function. The truncated orthologs were found to be enzymatically active toward xyloglucan, but with significantly reduced catalysis compared to the full-length proteins after 6 h of incubation with the substrate (*t* test; *P*-value = 0.0002 and 0.01 for *P. nicotianae* and *P. cactorum*, respectively). No reducing sugars were detected in the vector-only sample (n = 3, ±SD).

alignment variation from *P. sojae*_247788 (*SI Appendix*, Fig. S1). In contrast to *P. sojae*_247788, the orthologous proteins of *P. cactorum* and *P. nicotianae* demonstrated significantly higher xyloglucanase function (Fig. 3*B*). To investigate whether this function was specific to the presence or absence of the C-terminal extensions, truncated versions of the proteins (removing the C-terminal regions) were engineered and expressed in *S. cerevisiae*. The truncated orthologs in *P. nicotianae* and *P. cactorum* were found to be enzymatically active toward xyloglucan, but with significantly reduced catalysis compared to the full-length proteins after 6 h of incubation with the substrate (*t* test; *P*-value = 0.0002 and 0.01 for *P. nicotianae* and *P. cactorum*, respectively) (Fig. 3*B*). These data suggest that high-level xyloglucanase catalytic function within the 482953-247788 clade was, in part, a result of the acquisition of a highly disordered C-terminal extension, but that this high-level function has not been retained specifically in the *P. sojae*_247788 paralog, yet is present in the *P. cactorum* and *P. nicotianae* orthologous proteins.

The Functional *P. sojae*_559651 (PsXEG1) Xyloglucanase and Its Orthologs Are Predicted to Encode an Additional Substrate-Binding Site. As discussed above, *P. sojae*_559651 (PsXEG1) and its associated orthologs were shown to have two separate single amino acid insertions (*SI Appendix*, Fig. S1). Predictions of substrate-binding sites (active site residues) for GH12 paralogs were generated by 3DLigandSite (63). *P. sojae*_559651 (PsXEG1) was the only GH12 paralog in which an additional binding site was predicted; the amino acid residues for this second putative site were Gly55, Ala56, Ala57, Thr58, Val97, Phe205, Val206 (residue position numbers given for the protein sequence in the absence of its N-terminal signal peptide); Thr58 was the common residue predicted in both ligand sites for this paralog (*SI Appendix*, Fig. S5*A*). Orthologous proteins in *P. cactorum* and *P. nicotianae* were also predicted to encode this structurally inferred binding site. While all of the additional binding site residues were found to be conserved among the three orthologous protein sequences, for *P. cactorum*, Ala57 was not predicted to form part of its putative binding site. Interestingly, the two putative insertions coding for alanine (Ala, A) and Serine (Ser, S) conserved among the three orthologs, were important for defining the additional binding site prediction (*SI Appendix*, Fig. S1). In silico removal of both amino acids from the sequence prior to 3DLigandSite (63) analyses abolished prediction of the second binding site, while the "primary" binding site common among all GH12 paralogs remained intact. Of all the *P. sojae* GH12 paralog sequences, the two putative indels are present only in *P. sojae*_559651 (PsXEG1) and *P. sojae*_360375 (PsXLP1) (*SI Appendix*, Fig. S1), however, as the latter paralog is significantly truncated at the C terminus, a second putative binding site was not predicted for this protein. This paralog was not studied further, as it was investigated previously by Ma et al. (2017), with interesting insights into host–pathogen interactions (43).

To investigate the functional significance of the structurally inferred binding site, *P. sojae*_559651 (PsXEG1) and its ortholog identified from *P. cactorum* were engineered to delete both the A and S residues. The amended genes were heterologously expressed in *S. cerevisiae*, but we found no difference in xyloglucanase activity as monitored by the release of reducing sugars over 6 h of incubation (pH 7, 30 °C) (*SI Appendix*, Fig. S5*B*). Taken together, the data suggest an as yet unknown role, if any, for the structurally inferred additional binding site in *P. sojae*_559651 (PsXEG1).

***P. sojae*_482953, 260883, and 520599 Trigger Rapid Generation of ROS in *N. benthamiana*.** The generation of ROS in plants is a marker for stress and early immune responses (64). To investigate how plant immunity is modulated in response to each of the xyloglucanase variants, the eleven paralogs (and a vector-only negative control) were codon-optimized and expressed in *N. benthamiana* leaves via *Agrobacterium*-mediated infiltration, driven by a CaMV35S promotor (65) but maintaining the variant putative oomycete signal peptides (*SI Appendix*, Fig. S2*B*). Leaf discs were collected 48 h postinfiltration and incubated in water overnight. The generation of ROS was then measured from 0 to 60 min. ROS accumulation was detected within minutes of adding ROS assay reagent for leaf discs expressing *P. sojae*_482953, 260883, and 520599 paralogs, and remained high for the entire measurement period (Fig. 4*A*). These results demonstrate phylogenetically divergent *P. sojae* proteins (Fig. 1) with variant *P. sojae* signal peptides (*SI Appendix*, Fig. S2*B*) can trigger stress and early immune responses in plants. Total photon counts at 60 min suggest that the ROS response to *P. sojae*_482953 was significantly higher than that of *P. sojae*_260883 and 520599 (Fig. 4*D*). *P. sojae*_482953 represents one of the most active enzymes in this gene family under the yeast expression conditions tested (Fig. 2*A*); to further explore whether the ROS response triggered by this variant protein could be attributed to the presence of the C-terminal extension [which we demonstrated increases its xyloglucanase activity (Fig. 3*A*)], the *P. sojae*_482953 C-terminal tail only without a signal peptide was expressed in *N. benthamiana* leaves via the same *Agrobacterium*-mediated infiltration method. Interestingly, we find that the C-terminal extension alone triggers ROS generation, but significantly less than the full-length protein, and not statistically significant from the vector-only control after 60 min (*SI Appendix*, Fig. S2).

Intriguingly, ROS accumulation was also significantly enhanced by *P. sojae*_260883 and *P. sojae*_520599 proteins—these paralogs showed significantly reduced xyloglucanase activity compared with *P. sojae*_482953 (Fig. 2). This reduced ROS response could be a product of reduced enzymatic catalysis, reduced secretion efficiency from the variant oomycete signal peptides, less potent plant immune responses, or a combination of all three factors. Furthermore, gene expression of *P. sojae*_520599 is undetectable in all life-cycle stages sampled, while *P. sojae*_260883 shows limited expression during plant infection (at least at 3 d postinfection of soybean hypocotyls) in comparison with *P. sojae*_482953 (Fig. 1*B*).

Agrobacterium-mediated infiltration of *P. sojae*_247788 and 559651 (PsXEG1) also triggered ROS generation in *N. benthamiana*, but significantly reduced compared to *P. sojae*_482953, 260883, and 520599 proteins, and total photon counts at 60 min were not statistically significant from the vector-only control (Fig. 4*D*). Interestingly, ROS accumulation in response to these xyloglucanase paralogs peaked at around 20 min postinfiltration and declined by 60 min observation (Fig. 4*B*). This pattern was not observed in any of the variants that triggered the highest ROS accumulation (*P. sojae*_482953, 260883, and 520599). We did not find evidence of ROS accumulation as a result of infiltration with *P. sojae*_338064, 338074, 355355, 360375 (PsXLP1), 520248, 520924, and the *P. sojae*_247788 C-terminal extension only (Fig. 4*C*). The absence of ROS response across these proteins could again be due to variant enzymatic catalysis, reduced secretion efficiency from the variant oomycete signal peptides, less potent plant immune responses, or a combination of all these factors.

Bacterial-associated flg22 and fungal-associated chitin represent well-known pathogen-associated molecular patterns (PAMPs), both triggering ROS generation in planta (66–68). Our analyses reported above (Fig. 4), show that the *P. sojae* xyloglucanase paralogs can generate ROS responses. However, plant-pathogen relationships depend on numerous interactions. To explore whether the xyloglucanase variants could enhance or inhibit ROS generation from known elicitors of ROS, we expressed *P. sojae* paralogs in *N. benthamiana*, and applied flg22 or chitin exposure treatments. Interestingly, we find

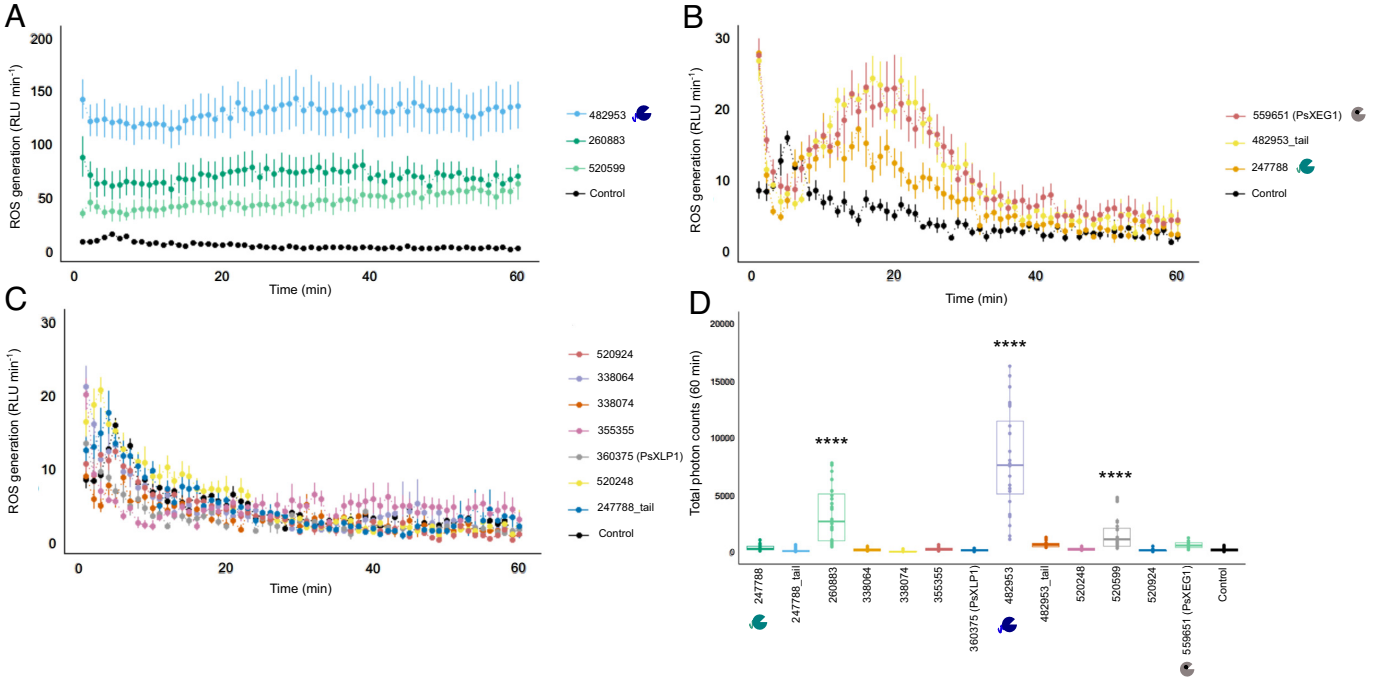

Fig. 4. *P. sojae* xyloglucanase paralogs induction of ROS responses in a model plant system. *P. sojae* xyloglucanase paralogs were codon-optimized for expression in plants and expressed in *N. benthamiana* leaves via *Agrobacterium*-mediated infiltration. The dynamics of ROS production in *N. benthamiana* plants was measured from 0 to 60 min. ROS production corresponds to the mean of eight samples (±SE). (*A*) High ROS accumulation was observed for *P. sojae*_482953, 260883, and 520599 paralogs. (*B*) *P. sojae*_559651, 482953 (tail only), and 247788 triggered a unique pattern of ROS that peaked at 20 min. (*C*) No ROS accumulation was detected for the remaining GH12 variants. (*D*) Total ROS production over 60 min. Horizontal bars represent the mean, and asterisks show *P*-value < 0.0001 using the *t* test. ROS production corresponds to the mean of 24 samples. Symbols indicate the *P. sojae* paralog with a structurally inferred additional binding site (gray), and paralogs with C-terminal tails 1 and 2 (blue and green cartoon symbols, respectively). Positive controls, i.e., elicitation of ROS response against flg22 during exposure to a vector backbone, are shown in *SI Appendix*, Fig. S3*A*.

that *P. sojae*_247788 and 260883 can significantly suppress flg22-triggered ROS generation in *N. benthamiana* (*SI Appendix*, Fig. S3*A*)—suggesting complex interactions, including both immune response promotion and suppression between xyloglucanase paralogs, within the dynamic plant environment. No significant chitin-triggered ROS generation was observed (*SI Appendix*, Fig. S3*B*).

***P. sojae* GH12 Paralogs Cleave the Same Glycosidic Linkages of the Xyloglucan Backbone, but Show Differences in Their Oligosaccharide Products.** Xyloglucan oligosaccharides released by *P. sojae*, *P. nicotianae*, and *P. cactorum* xyloglucanases (or a vector-only control) upon digestion of 1% (w/v) xyloglucan (72 h, pH 7, 30 °C), were analyzed by mass spectrometry (MALDI) (analysis carried out by Wyatt Analytical Ltd). Four peaks of interest in the mass spectra were observed across the samples containing functional enzyme variants; ions with *m/z* of ~1,085, 1,247, 1,409, and 1,571—putatively corresponding to the xyloglucan oligosaccharides XXXG, XXLG (or XLXG), XLLG, and XLFG, respectively (24). These xyloglucan oligosaccharide compounds were not observed in the vector-only controls—consistent with the observed xyloglucanase activity of the functional enzyme variants (Fig. 2). The relative intensities of oligosaccharide species identified at m/z ~1,085, 1,247, 1,409, and 1,571 were compared by calculating the ratios between the areas under the peaks.

Fig. 5 provides strong evidence that the composition of oligosaccharides varies significantly between active *P. sojae* xyloglucanase paralogs (Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001). Goodness-of-fit testing indicates that the Dirichlet distribution provides an adequate model for these data (*P*-value = 0.48). We suggest that putative differences in enzymatic output across the wider GH12 family could act to increase the efficiency of substrate breakdown during *P. sojae* growth and infection (Fig. 5). We see virtually undetectable levels of XLFG for *P. sojae*_482953, *P. sojae*_559651 (PsXEG1), and *P. sojae*_338074, representing the two most active xyloglucanase variants under the conditions we tested [i.e., *P. sojae*_482953 and *P. sojae*_559651 (PsXEG1)], with *P. sojae*_338074 also showing significant enzymatic activity at 6 h (Fig. 2*A*)—suggesting that the individual enzymatic activities of these variants do not act to generate this oligosaccharide product. We also observed an increased output of XLFG for *P. sojae*_355355, and this paralog is expressed during all three *P. sojae* lifecycle stages (it is notably among the most highly expressed of the GH12 paralogs in the cyst and infection stages; Fig. 1*B*). Our results also demonstrate increased output of XXXG for *P. sojae*_260883, and this paralog is among the most highly expressed of the GH12 paralogs in the cyst stage (Fig. 1*B*), as well as triggering significant ROS accumulation in *N. benthamiana* (Fig. 4*A*). Differences in enzymatic output between the enzyme variants across different stages of the *P. sojae* lifecycle could result in more effective breakdown of xyloglucan, but could also act to diversify the types of xyloglucan oligosaccharides that are released and therefore perceived as DAMPs by the plant host.

We also tested whether the composition of oligosaccharides varied between full-length and truncated xyloglucanase variants for each ortholog. For C-terminal extension 1 (*P. sojae*_482953; Fig. 6*A*), the analysis indicated a significant difference in composition of oligosaccharides produced by the full-length and truncated *P. nicotianae* variants [Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001], but no significant difference between *P. sojae* (*P*-value = 0.098) or *P. cactorum* (*P*-value = 0.104) variants. For C-terminal extension 2 (*P. sojae*_247788; Fig. 6*B*), the analysis indicated a significant difference in composition of oligosaccharides produced by full-length and truncated *P. nicotianae* variants [Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001], but no significant difference between *P. cactorum* variants (*P*-value = 0.392) variants. We find

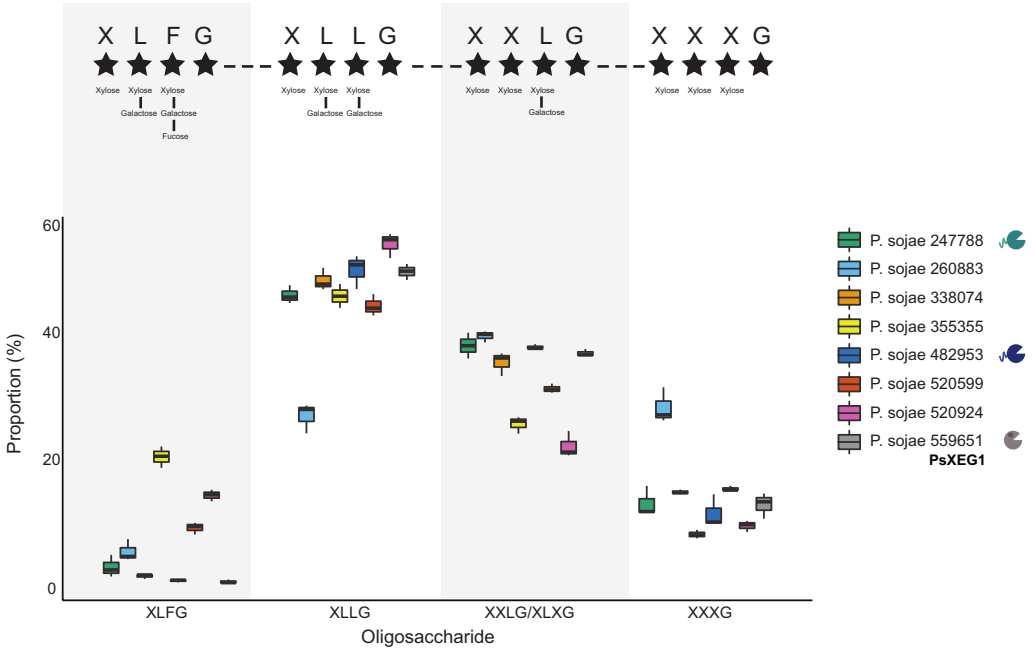

Fig. 5. *P. sojae* xyloglucanase paralogs produce variant oligosaccharide break-down products. *P. sojae* xyloglucanase paralogs secreted into *S. cerevisiae* culture supernatants were incubated with 1% (w/v) xyloglucan at 30 °C, pH 7, for 72 h. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS) spectra confirmed the release of xyloglucan oligosaccharides by eight of the 11 variants (corresponding to the functional paralogs identified in Fig. 2). Four peaks of interest were observed; ions with *m/z* of ~1,085, 1,247, 1,409, and 1,571—putatively corresponding to the oligosaccharides XXXG, XXLG (or XLXG), XLLG, and XLFG respectively (24). The *Top* panel indicates the side chain residues present in each of the oligosaccharide variant types, as described by Fry et al., 1993 (24). The relative intensities of species identified at these peaks were compared by calculating the ratios between the areas under the peaks, to probe putative differences in preferential binding of the xyloglucan backbone. We find strong evidence that the composition of oligosaccharides varies significantly between the *P. sojae* xyloglucanase paralogs [Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001]. For example, we see almost absence of the XLFG product for three enzymes. Goodness-of-fit testing indicates that the Dirichlet distribution provides an adequate model for these data (*P*-value = 0.48).

no evidence of significantly altered oligosaccharide profiles when comparing *P. sojae*_559651 (PsXEG1) with and without A and S residues (*SI Appendix*, Fig. S5*C*).

We also tested whether individual xyloglucan oligosaccharides could elicit a ROS response in *N. benthamiana*. The oligosaccharides XLFG, XLLG, and XXXG (24) were infiltrated into the leaves of *N. benthamiana* and leaf discs were subject to a ROS assay, alongside a negative control (H_2O) and a positive control (flg22). We found no evidence of ROS accumulation as a result of infiltration with any of the xyloglucan oligosaccharides tested (*SI Appendix*, Fig. S4), consistent with previous data that confirm xyloglucan oligosaccharides do not trigger a ROS response, but rather a broad range of alternative immune responses have been reported (25).

Discussion

The genomes of necrotrophic and hemibiotrophic plant pathogens possess many genes predicted to encode secreted degradative enzymes (26, 27). Many of these enzymes are thought to function in breakdown of the plant cell wall and additional plant tissues, which are both defensive structures and rich in valuable saccharides (17, 18). HGT has been shown to be a key factor in the evolution of gene families predicted to function in breakdown of plant tissues (28–33). Many of these HGT gene families demonstrate evidence of multiple rounds of gene duplication post-transfer, pointing toward a process of repeat selection for retention of duplicated genes, which may subsequently drive transcriptional amplification (gene dosage) and/or neofunctionalization of gene paralogs. To explore this phenomenon further, we investigated the evolution of protein function within a HGT gene family (GH12) in *P. sojae*. We were able to demonstrate that this gene family encodes eight paralogs with consistently detectable xyloglucanase function, each with variant transcriptional profiles. We show that some functional paralogs have diversified additional structural features important for xyloglucanase function, as well as unique ROS accumulation profiles in *N. benthamiana* for a subset of the *P. sojae* paralogs. We also show that xyloglucanase function among the wider *P. sojae* GH12 family results in significantly varied enzyme digestion products. Taken together, these results suggest that gene duplication and paralog evolution has played an important role not only in the diversification of xyloglucanase function, but in the alteration of host perception during pathogen attack.

*P. sojae*_482953 and *P. sojae*_247788 encode significantly disordered C-terminal extensions; truncation of *P. sojae*_482953, and orthologous proteins in *P. nicotianae* and *P. cactorum* reduces enzymatic activity upon xyloglucan—consistent with a hypothesis that the C-terminal extensions play an important role in maintaining interactions with the enzyme substrate (therefore optimizing the encoded enzyme activity at the conserved catalytic sites). Interestingly, Li et al. (2014) describe a C-terminal proline-rich sequence of xylanase XynA, and its removal also negatively affects the protein function (69). In contrast, Wen et al. (2005) describe a truncated glucanase that displays improved enzymatic activity compared to its full-length counterpart (70), suggesting that evolution of a C-terminal extension can both positively and negatively affect enzymatic function, depending on their interactions with putative substrates. We find that *P. sojae*_247788 (full-length) has minimal xyloglucanase function, although interestingly, this paralog has significantly reduced expression at all of the life-cycle stages tested previously [FungiDB; (55, 56)] (Fig. 1*B*). Conversely, full-length orthologs in *P. cactorum* and *P. nicotianae* display increased xyloglucanase activity, and truncation of the C-terminal extensions reduces the rate at which both proteins can digest xyloglucan. These data suggest that addition of

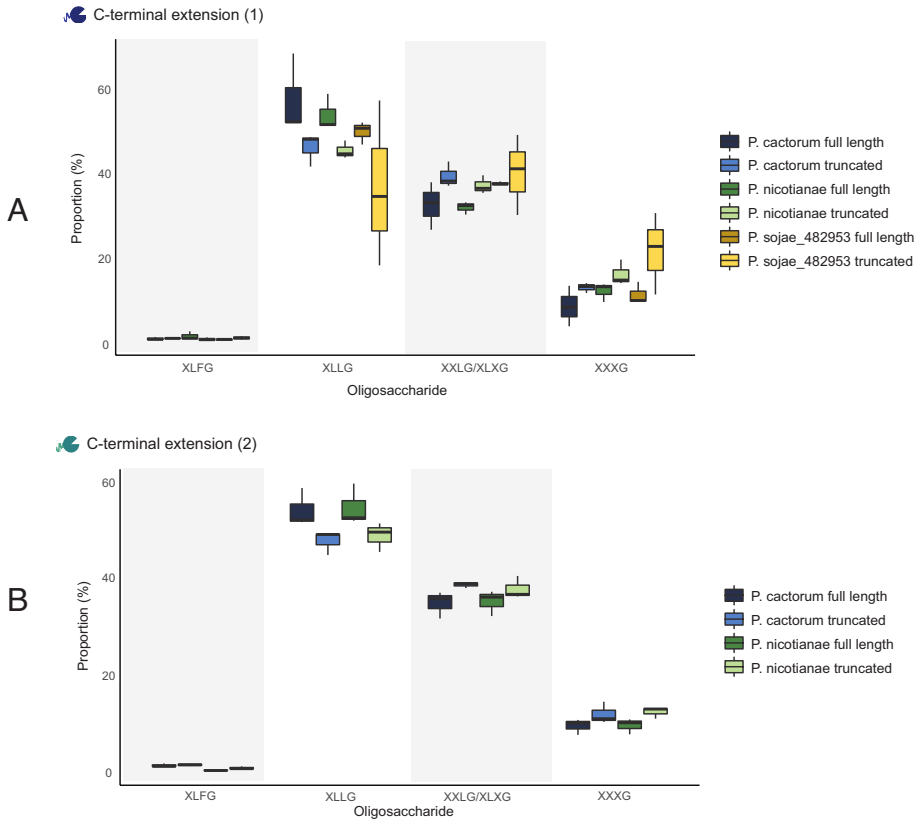

Fig. 6. C-terminal tail extensions produce variant oligosaccharide break-down products in only two of the five protein variants tested. We tested whether the composition of oligosaccharides varied between the full-length and truncated xyloglucanase variants for each ortholog. Xyloglucanase variants secreted into *S. cerevisiae* culture supernatants were incubated with 1% (w/v) xyloglucan at 30 °C, pH 7, for 72 h. MALDI-MS spectra confirmed the release of xyloglucan oligosaccharides. Four peaks of interest were observed; ions with *m/z* of ~1,085, 1,247, 1,409, and 1,571—putatively corresponding to the oligosaccharides XXXG, XXLG (or XLXG), XLLG, and XLFG, respectively (24). The relative intensities of species identified at these peaks were compared by calculating the ratios between the areas under the peaks, to probe putative differences in preferential binding of the xyloglucan backbone. (*A*) For C-terminal extension 1 (*P. sojae*_482953), we saw a significant difference in the composition of oligosaccharides produced by the full-length and truncated *P. nicotianae* variants [Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001], but there was no significant difference between each of the *P. sojae* (*P*-value = 0.098) or *P. cactorum* (*P*-value = 0.104) variants. (*B*) For C-terminal extension 2 (*P. sojae*_247788), we saw a significant difference in the composition of oligosaccharides produced by the full-length and truncated *P. nicotianae* variants (Dirichlet likelihood ratio test (randomization approach); *P*-value < 0.001), but there was no significant difference between the *P. cactorum* variants (*P*-value = 0.392). The MS spectra for *P. sojae_247788* (truncated) were noisy, so this paralog was omitted from (*B*) and our analysis.

these C-terminal extensions has further amended the function of the HGT-derived xyloglucanase post-gene duplication events.

Previous work by Claverie et al. (2018) demonstrates that mixed xyloglucan oligosaccharides do not trigger ROS accumulation (25). To test this further, we exposed three high-purity xyloglucan oligosaccharides identified in our mass spectrometry analyses (Fig. 5) to *N. benthamiana* and assayed for a ROS response. These analyses demonstrate no detectable ROS response, consistent with the work of Claverie and coauthors. We therefore hypothesize that ROS accumulation observed in response to the xyloglucanase variants could be triggered by recognition of the proteins rather than the products of enzyme digestion. Consistent with this hypothesis our data demonstrate differential responses; *P. sojae*_482953, 260883, and 520599 resulted in significant accumulation of ROS, while *P. sojae*_247788 and 260883 suppressed the ROS response elicited by flg22 (Fig. 4 and *SI Appendix*, Fig. S3*A*). These contrasting results suggest that at least four of *P. sojae* xyloglucanase proteins have effects on the *N. benthamiana* immune response. This is an important result because it also demonstrates that different paralogs have different effects on the plant immune response.

In previous work by Ma et al. (2015), the authors demonstrate that *P. sojae*_559651 (PsXEG1), 482953, 260883, and 247788 induce cell death in *N. benthamiana* (44)—consistently, we detected ROS accumulation as a result of expression of these variants, albeit at remarkedly different levels between paralogs—a result that could be influenced by different protein characteristics, different rates of enzymatic catalysis, or different secretion efficiencies in our *N. benthamiana* expression system. Interestingly, *P. sojae*_520599 was not demonstrated to induce cell death in previous work (44), but our results suggest infiltration of this protein into *N. benthamiana* leaves results in the third-highest accumulation of ROS among the xyloglucanase variants in this gene family. According to publicly available RNA-seq data for *P. sojae*, there is no evidence that *P. sojae*_520599 is transcriptionally active during plant infection (Fig. 1*B*, although, we acknowledge that the transcriptome data are limited to just one time point during infection)—but it is interesting to consider whether the reduced expression of this gene is a response to the host–pathogen arms race to limit host immune perception. Further focused work on this xyloglucanase paralog would be useful to uncover how the evolution of this protein has been shaped by host immune responses over evolutionary time.

Investigation of xyloglucan oligosaccharides released following xyloglucan digestion suggests that enzymatic activity of *P. sojae* GH12 paralogs leads to diversified enzymatic output. We hypothesize that this example of neofunctionalization can offer a greater efficacy of in vivo xyloglucan digestion. We also suggest that the variation in oligosaccharides released as a product of the function of active GH12 paralogs may contribute to the diversification of the DAMPs released during pathogen invasion. Previous work by

Claverie et al. (2018) confirms that xyloglucan oligosaccharides are recognized as DAMPs and trigger a broad range of ROS-independent immune responses (25) (although we do not know whether different xyloglucan oligosaccharides promote the same response). It is therefore possible that duplication and neofunctionalization of xyloglucanase output would require the host immune system to maintain a high sensitivity of detection for one or more specific DAMP signatures among a diversity of alternative carbohydrate breakdown products present in dynamic and fluctuating relative concentrations. Further work to disentangle host responses to individual xyloglucan oligosaccharides would help to fully understand the relevance, if any, of the diversified enzymatic output we observed in our study (Fig. 5). Due to structural variation of xyloglucan between plant species, further work to compare enzyme function against different xyloglucans will be required to understand how polysaccharide structure relates to the diversity of enzyme responses we observed in our study.

The Red Queen hypothesis (inspired by Lewis Carroll's 1871 novel, *Through the Looking-Glass*) is useful for understanding host–pathogen coevolutionary dynamics (45, 46) and predicts that evolution will act to generate patterns of diversification on any molecular signature that activates a host immune response (e.g., ref. 47). Consequently, the pattern of gene duplication, loss, deactivation, and neofunctionalization seen in the *P. sojae* GH12 xyloglucanase paralog family could be a factor in the host–pathogen arms race. Specifically, we present a variant paralog with a C-terminal extension that increases enzymatic activity, as well as demonstrate diverse ROS accumulation patterns in planta, and diverse profiles of enzymatic breakdown products among the members of this HGT xyloglucanase family. We therefore suggest that the host immune system must perceive and respond to fluctuations in multiple protein signatures, enzymatic catalysis rates and, potentially, multiple DAMP signatures. If the host fails to recognize the function or effects of one paralog, a selective sweep would quickly favor a variant strain, utilizing a paralog that produced a profile that was not detected, allowing immune-evasion and therefore selection for that variant. Consistent with this hypothesis, the gene family shows multiple paralogs that have reduced detectable xyloglucanase function, paralog forms which trigger minimal or reduced ROS accumulation, and others which show near-absence of some classes of oligosaccharide output—a possible product of pathogen protein evolution with the selected outcome of escaping host immune detection. If, and when, the plant immune system catches up with a pathogen's amended protein signature, a variant set of paralogs with an altered signature could then be selected for prominence, allowing improved evasion of the plant immune system and resulting in fixation of additional gene duplication forms with neofunctionalized functions. Collectively, we argue the data presented here and in Ma et al. (43, 44) demonstrate that the *P. sojae* GH12 xyloglucanase paralog families show this evolutionary dynamic. Taken together, these results contribute to a better understanding of the functional differences between gene duplications post-HGT, giving us improved insight into the dynamics of phytopathogenicity, as well as a better understanding of the consequences of gene flow in phytopathogenic oomycetes.

Materials and Methods

Phylogenetic Analyses. Candidate homologs (paralogs) of *P. sojae* GH12 (Pfam: PF01670, InterPro: IPR002594) were collated by performing HMM searches against Ensembl genomes (restricted by taxonomy to *P. sojae* NCBI taxids 67593 and 1094619) with default parameters (48), using the raw profile HMM training set for the putative protein family [Pfam: (49)]. Additional oomycetes were selected for sampling based on the availability of complete genome sequence datasets and representing a diversity of hemibiotrophic forms. We also included the obligate biotrophic pathogen *H. arabidopsidis* for comparison. It is likely that further oomycete genome sampling would further complicate the pattern of gene duplication identified. A phylogenetic tree was constructed to investigate the evolutionary relationships between putative *P. sojae* GH12 paralogs and other oomycetes. GH12 protein homologs were identified from a selection of eukaryotic (with sampling subsequently focusing on oomycetes and fungi) and prokaryotic genomes using BLASTp (71); from these hits, a multiple sequence protein alignment was constructed and aligned using automated methods in Seaview (72) using MUSCLE (73), which was then edited and masked manually. GenBank was checked for additional gene sampling of oomycetes not included in the genome-by-genome survey, and additional prokaryote and eukaryote–including fungal–genes. A preliminary phylogenetic tree was calculated using PhyML (74) to confirm all *P. sojae* GH12 paralogs sampled were part of the same oomycete monophyletic clade and predicted to be a HGT from fungi (31, 33); this was then checked manually, enabling the removal of distantly related fungal and prokaryotic outgroups and partial/incomplete "in group" sequences indicating poor gene models. The alignment was subject to an additional round of manual editing and masking, resulting in an alignment with 161 sequences and 211 amino acid sites. A second PhyML tree was calculated and checked, and the final ML tree was constructed with IQ-Tree v2.0.3 with ModelFinder (52) identifying the WAG+R5 model. The tree was subject to nonparametric bootstrap with 200 pseudoreplicates (54) and rooted with a fungal outgroup (31, 33). The final tree was visualized with iTOL (53).

Comparisons of Putative Oomycete GH12 Protein Structures. *P. sojae* GH12 protein sequences were aligned using Clustal Omega (75, 76), and visualized in Jalview2 (77). Three-dimensional structures were obtained with Phyre2 (Protein Homology/analogY Recognition Engine v2.0 (60) or AlphaFold (61, 62), using protein sequences without their predicted N-terminal signal peptide [determined using SignalP-6.0 (78)]. Predictions of carbohydrate-binding sites for putative paralogous proteins were generated by 3DLigandSite (date searched: June/2021); the server was used to identify high-scoring homologous protein structures with ligands bound (by the comparative MAMMOTH score) (63).

Identification of Transcription Responses of the GH12 Gene Family in *P. sojae*. FungiDB contains comparative RNA-seq data for *P. sojae* ensemble transcriptome samples from life-cycle stages: mycelial, cyst, and 3 d postinfection (soybean hypocotyls infected with *P. sojae* strain P6497) (55, 56). Relative transcription of the GH12 paralogs was identified as transcript levels of FPKMs [sequencing depth and gene depth normalized from paired-end RNA-seq data; FungiDB; (55, 56)]. RNA-seq data for *P. sojae* were additionally used to manually check GH12 gene models prior to the synthesis of genes for experimental study.

Gene Synthesis of GH12 Genes from *P. sojae* and Relatives. We identified 11 *P. sojae*-specific GH12 paralogs (Fig. 1). All 11 *P. sojae* paralogs were codon-optimized for translation in *S. cerevisiae* (http://www.genscript.com/tools/rare-codon-analysis). The gene sequences were synthesized by Synbio Technologies into the vector p426-GPD (American Type Culture Collection (ATCC) 87361) between *BamHI* (GGATCC) and *HindIII* (AAGCTT) restriction sites. *P. sojae* N-terminal signal peptide sequences were replaced with *S. cerevisiae* MFA (57) (*SI Appendix*, Table S1), in order to maximize heterologous protein secretion in yeast.

Assays for GH12 Xyloglucanase Function. To compare xyloglucanase function, we conducted activity assays against a commercially available xyloglucan [Tamarind; backbone of β-1,4-glucan; most substituted (Megazyme)]. Yeast transformants were screened for released reducing sugars using DNS reagent as follows: Recombinant *S. cerevisiae* strains (in biological triplicate and including a p426-GPD vector-only control strain) were cultured in 20 mL synthetic complete medium (1% glucose) lacking uracil (SCM-URA) for 7 d at 30 °C with 180 rpm shaking. Culture supernatants were removed by centrifugation at 4 °C, concentrated 10× (Corning Spin-X UF concentrator; 10,000 kDa MWCO), and kept on ice. Total protein concentrations were measured using a Qubit Fluorometer and Thermo Scientific reagents. Concentrated supernatants were matched to 75 μg/mL and incubated with 1% (w/v) xyloglucan in 50 mM citrate buffer (pH 7) at 30 °C with 180 rpm shaking. At each time point sampled, released reducing sugars

were measured by removing 60 μL of each sample into a sterile 1.5 mL tube, and adding 60 μL of DNS reagent [1% (w/v) DNS, 1 M potassium sodium tartrate, and 400 mM sodium hydroxide], followed by incubation at 95 °C for 5 min to allow color development. Samples were transferred to a sterile 96-well plate, and the absorbance was measured at 544 nm using an absorbance microplate reader (CLARIOstar; BMG LABTECH).

For the Congo red assays (*SI Appendix*, Fig. S6*A*), 10 μL of the concentrated supernatants at 75 μg/mL total protein was spotted onto Synthetic Complete Medium minus Uracil agar (SCM-URA) containing 0.2% (w/v) xyloglucan [with 2% (w/v) glucose as an additional carbon source]. Agar plates were incubated at 30 °C for 24 h (static), and the remaining intact polysaccharide on the plates was stained with 0.2% (w/v) Congo red (Sigma Aldrich) for 30 min at room temperature, and destained with 1 M sodium chloride for 30 min (with shaking) at room temperature (79). Extracellular enzyme activity was indicated by a clearing or "halo" around the spots containing secreted functional enzyme.

For the cell culture–based Congo red assays (*SI Appendix*, Fig. S6*B*), we followed an identical protocol as above, except 10 μL of *S. cerevisiae* cell culture (i.e. prior to centrifugation and collecting the supernatant) was spotted onto the SCM-URA [0.2% (w/v) xyloglucan] agar plates.

All data analysis was performed using R version 4.0.3. DNS assay data (Fig. 2) were corrected using the initial (0 h) OD reading for each sample, to account for residual sugars present in concentrated supernatant samples; these data were then analyzed using a Dunnett test against the vector-only control, using package DescTools (80). DNS assay data (Fig. 3 and *SI Appendix*, Fig. S5*B*) are displayed as OD readings at each timepoint; 6 h data were then analyzed using a Student's *t* test against the vector-only control (two-tailed, two-sample equal variance) (81).

Detection of ROS in *N. benthamiana* Plants in Response to Xyloglucanase Expression or Xyloglucan Constituent Oligosaccharides. GH12 paralogs were codon-optimized for translation in *N. benthamiana* plants with their native *P. sojae* signal peptide sequences (see *SI Appendix*, Fig. S2*B* for alignment of N-terminal regions), and the gene sequences were synthesized by Genewiz and Twist Bioscience. Synthesized DNA fragments were cloned into vector pGWB514 (Addgene 74856) (82) by In-Fusion cloning (Takara Bio)–expression was driven by CAMV35S promoter (65). Verified plasmids were then introduced into *Agrobacterium* strain AGL1 by a freeze–thaw method (83). Four-week-old *N. benthamiana* leaves were infiltrated with *Agrobacterium* strain AGL1 (OD_{600} = 0.4) containing the corresponding GH12 genes. Eight leaf discs from each infiltrated vector were collected at 48 h post infiltration. Leaf discs were incubated in water in white 96-well plates (Greiner Bio-One 655075) overnight. Water was then replaced with ROS assay reagent containing 100 μM Luminol (Merck 123072) and 20 μg/mL horseradish peroxidase (Merck P6782). ROS generation was measured at 0 to 60 min using a High-Resolution Photon Counting System (Photek HRPCS218). For the PAMP treatment assay, 1 μM flg22 or 1 mg/mL chitin was added to the ROS assay reagent.

For xyloglucan oligosaccharide treatment assays, the oligosaccharides XLFG, XLLG, and XXXG were sourced from BOC Sciences (XXLG/XLXG was unable to be synthesized; therefore, it was not possible to test this particular oligosaccharide variant in this assay). All xyloglucan oligosaccharides were prepared at 1 mg/mL in H_2O, and infiltrated into the leaves of *N. benthamiana*. Eight leaf discs for each infiltrated treatment were collected at 48 h post infiltration, and ROS assays were carried out as described above.

Mass Spectrometry Analysis of Xyloglucan Digestion Products. Mass spectrometry was utilized to investigate the differences in degradation products released from xyloglucan breakdown by the xyloglucanase variants. Yeast transformants were prepared for MS analysis as follows: recombinant *S. cerevisiae* strains (in biological triplicate and including a p426-GPD vector–only control strain) were cultured in 20 mL SCM-URA (1% glucose) for 7 d at 30 °C with 180 rpm shaking; the supernatants were removed by centrifugation and concentrated to 10×, as described above. Concentrated supernatants (matched to 75 μg/mL) were incubated with 1% (w/v) xyloglucan in 50 mM citrate buffer, pH 7 for 72 h at 30 °C with 180 rpm shaking. At 0 and 72 h, 60 μL of each sample was removed into a sterile 1.5 mL tube, and the reactions terminated by heating at 95 °C for 5 min. The samples were air-dried using a centrifugal vacuum concentrator. MALDI data were acquired in positive-reflectron mode (Bruker ultrafleXtreme spectrometer); all sample preparation, optimization, and sample analyses were carried out by Wyatt Analytical Ltd.

Mass spectrometry data for paralog oligosaccharide diversity were analyzed using a Dirichlet likelihood ratio test using R code published previously (84). A randomization procedure was used to determine *P*-values (5,000 random samples), and goodness-of-fit testing was used to test whether the Dirichlet distribution provided an adequate model for this dataset, using 10,000 simulations (84).

Accession Numbers. Sequence data can be found in the NCBI protein database under the following accession numbers: *P. sojae*_482953 (PHYSODRAFT_482953; EGZ25667.1), *P. sojae*_247788 (PHYSODRAFT_247788; EGZ25668.1), *P. sojae*_559651(PHYSODRAFT_559651; **PsXEG1**; EGZ16757.1), *P.sojae*_355355(PHYSODRAFT_355355; EGZ11358.1), *P. sojae*_260883 (PHYSODRAFT_260883; EGZ11355.1), *P. sojae*_520924 (PHYSODRAFT_520924; EGZ11360.1), *P. sojae*_520599 (PHYSODRAFT_520599; EGZ11361.1), *P. sojae*_360375(PHYSODRAFT_360375; **PsXLP1**; EGZ16758.1), *P.sojae*_338074(PHYSODRAFT_338074; EGZ11362.1), *P.sojae*_338064(PHYSODRAFT_338064; EGZ11350.1), *P. sojae*_520248 (PHYSODRAFT_520248; EGZ11363.1), *P. nicotianae*_extension1 (AM587_10010125; KUF85461.1), *P. nicotianae*_extension2 (AM587_10010124; KUF85460.1), *P.nicotianae*_binding (AM587_10012326; KUF76577.1), *P. cactorum*_extension1 (PC111_g4312; KAG2838309.1), *P. cactorum*_extension2 (Pcac1_g9740; KAG2780169.1), and *P cactorum*_binding (Pcac_g2817; KAG2788088.1).

Data, Materials, and Software Availability. The sequence alignments and underlying data for all plots are available at https://dx.doi.org/10.6084/m9.figshare.21524886 (85).

ACKNOWLEDGMENTS. V.A. and this research was supported by a Philip Leverhulme Award (PLP-2014-147) to T.A.R. T.A.R. is supported by a Royal Society University Research Fellowship (URF/R/191005). X.Y. is supported by an award from The Gatsby Charitable Foundation and Biological Sciences Research Council BBS/E/J/000PR9797 to N.J.T. Y.F. and J.H. were supported by NIH/NIAID R01 AI039115-26.

Author affiliations: [a]Department of Biology, University of Oxford, Oxford OX1 3SZ, United Kingdom; [b]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710; and [c]The Sainsbury Laboratory, University of East Anglia, Norwich Research Park NR4 7UH, United Kingdom

Author contributions: V.A., D.S.M., X.Y., Y.F., J.H., N.J.T., and T.A.R. designed research; V.A., D.S.M., Y.F., X.Y., G.L., and T.A.R. performed research; V.A., D.S.M., X.Y., G.L., and T.A.R. analyzed data; and V.A., D.S.M., J.H., N.J.T., and T.A.R. wrote the paper.

1. G. Van der Auwera *et al.*, The phylogeny of the Hyphochytriomycota as deduced from ribosomal RNA sequences of *Hyphochytrium catenoides*. *Mol. Biol. Evol.* **12**, 671–678 (1995).
2. S. L. Baldauf, The deep roots of eukaryotes. *Science* **300**, 1703–1706 (2003).
3. T. Cavalier-Smith, E.E.-Y. Chao, Phylogeny and megasystematics of phagotrophic heterokonts (Kingdom Chromista). *J. Mol. Evol.* **62**, 388–420 (2006).
4. I. Riisberg *et al.*, Seven gene phylogeny of heterokonts. *Protist* **160**, 191–204 (2009).
5. G. W. Beakes, M. Thines, "Hyphochytriomycota and Oomycota" in *Handbook of the Protists*, J. M. Archibald, A. G. B. Simpson, C. H. Slamovits, Eds. (Springer International Publishing, 2017), pp. 435–505.
6. T. A. Richards, N. J. Talbot, Horizontal gene transfer in osmotrophs: playing with public goods. *Nat. Rev. Microbiol.* **11**, 720–727 (2013).
7. G. C. Ainsworth, Ainsworth and Bisby's Dictionary of the Fungi, 5th edition (Commonwealth Mycological Institute, Kew, 1961).
8. H. Förster, M. O. Coffey, H. Elwood, M. L. Sogin, Sequence analysis of the small subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution. *Mycologia* **82**, 306–312 (1990).
9. M. C. Leclerc, J. Guillot, M. Deville, Taxonomic and phylogenetic analysis of *Saprolegniaceae* (Oomycetes) inferred from LSU rDNA and ITS sequence comparisons. *Antonie Van Leeuwenhoek* **77**, 369–377 (2000).
10. D. S. S. Hudspeth, S. A. Nadler, M. E. S. Hudspeth, A COX2 molecular phylogeny of the Peronosporomycetes. *Mycologia* **92**, 674–684 (2000).
11. D. S. S. Hudspeth, D. Stenger, M. E. S. Hudspeth, A cox2 phylogenetic hypothesis for the downy mildews and white rusts. *Fungal Divers.* **13**, 47–57 (2003).
12. M. Thines, Characterisation and phylogeny of repeated elements giving rise to exceptional length of ITS2 in several downy mildew genera (Peronosporaceae). *Fungal Genet. Biol.* **44**, 199–207 (2007).
13. C. G. P. McCarthy, D. A. Fitzpatrick, Phylogenomic reconstruction of the oomycete phylogeny derived from 37 genomes. *mSphere* **2**, e00095-17 (2017).
14. M. Latijnhouwers, P. J. G. M. de Wit, F. Govers, Oomycetes and fungi: Similar weaponry to attack plants. *Trends Microbiol.* **11**, 462–469 (2003).
15. N. P. Money, C. M. Davis, J. P. Ravishankar, Biomechanical evidence for convergent evolution of the invasive growth process among fungi and oomycete water molds. *Fungal Genet. Biol.* **41**, 872–876 (2004).
16. L. P. N. M. Kroon, H. Brouwer, A. W. A. M. de Cock, F. Govers, The genus phytophthora anno 2012. *Phytopathology* **102**, 348–364 (2012).
17. M. McNeil, A. Darvill, S. Fry, P. Albersheim, Structure and function of the primary cell walls of plants. *Annu. Rev. Biochem.* **53**, 625–63 (1984).

18. T. M. Schindler, The new view of the primary cell wall. *Z. Für Pflanzenernähr. Bodenkd.* **161**, 499–508 (1998).
19. W. S. York, H. van Halbeek, A. G. Darvill, P. Albersheim, Structural analysis of xyloglucan oligosaccharides by 1H-n.m.r. spectroscopy and fast-atom-bombardment mass spectrometry. *Carbohydr. Res.* **200**, 9–31 (1990).
20. W. S. York, V. S. Kumar Kolli, R. Orlando, P. Albersheim, A. G. Darvill, The structures of arabinoxyloglucans produced by solanaceous plants. *Carbohydr. Res.* **285**, 99–128 (1996).
21. J. P. Vincken, W. S. York, G. Beldman, A. G. Voragen, Two general branching patterns of xyloglucan, XXXG and XXGG. *Plant Physiol.* **114**, 9–13 (1997).
22. M. Hoffman *et al.*, Structural analysis of xyloglucans in the primary cell walls of plants in the subclass *Asteridae*. *Carbohydr. Res.* **340**, 1826–1840 (2005).
23. M. J. Peña, A. G. Darvill, S. Eberhard, W. S. York, M. A. O'Neill, Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. *Glycobiology* **18**, 891–904 (2008).
24. S. C. Fry *et al.*, An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiol. Plant.* **89**, 1–3 (1993).
25. J. Claverie *et al.*, The cell wall-derived xyloglucan is a new DAMP triggering plant immunity in *Vitis vinifera* and *Arabidopsis thaliana*. *Front. Plant Sci.* **9**, 1725 (2018).
26. C. A. Lévesque *et al.*, Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **11**, R73 (2010).
27. M. M. Zerillo *et al.*, Carbohydrate-active enzymes in pythium and their role in plant cell wall and storage polysaccharide degradation. *PLoS One* **8**, e72572 (2013).
28. T. A. Torto, L. Rauser, S. Kamoun, The pipg1 gene of the oomycete *Phytophthora infestans* encodes a fungal-like endopolygalacturonase. *Curr. Genet.* **40**, 385–390 (2002).
29. T. A. Richards, J. B. Dacks, J. M. Jenkinson, C. R. Thornton, N. J. Talbot, Evolution of filamentous plant pathogens: Gene exchange across eukaryotic kingdoms. *Curr. Biol.* **16**, 1857–1864 (2006).
30. L. Belbahri, G. Calmin, F. Mauch, J. O. Andersson, Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. *Gene* **408**, 1–8 (2008).
31. T. A. Richards *et al.*, Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15258–15263 (2011).
32. I. Misner, N. Blouin, G. Leonard, T. A. Richards, C. E. Lane, The secreted proteins of *Achlya hypogyna* and *Thraustotheca clavata* identify the ancestral oomycete secretome and reveal gene acquisitions by horizontal gene transfer. *Genome Biol. Evol.* **7**, 120–135 (2015).
33. F. Savory, G. Leonard, T. A. Richards, The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog.* **11**, e1004805 (2015).
34. B. A. Kronmiller *et al.*, Comparative genomic analysis of 31 phytophthora genomes reveals genome plasticity and horizontal gene transfer. *Mol. Plant Microbe Interact. MPMI* **36**, 26–46 (2023).
35. D. Soanes, T. Richards, Horizontal gene transfer in eukaryotic plant pathogens. *Annu. Rev. Phytopathol.* **52**, 583–614 (2014).
36. R. P. Hirt, C. Alsmark, T. M. Embley, Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. *Curr. Opin. Microbiol.* **23**, 155–162 (2015).
37. F. R. Savory, D. S. Milner, D. C. Miles, T. A. Richards, Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol. Biol. Evol.* **35**, 1887–1900 (2018).
38. B. N. Adhikari *et al.*, Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS One* **8**, e75072 (2013).
39. J. McGowan, D. A. Fitzpatrick, Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. *mSphere* **2**, e00408–17 (2017).
40. S. Ohno, *Evolution by Gene Duplication* (Springer Science & Business Media, 2013).
41. A. Stoltzfus, On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
42. A. Force *et al.*, Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
43. Z. Ma *et al.*, A paralogous decoy protects *Phytophthora sojae* apoplastic effector PsXEG1 from a host inhibitor. *Science* **355**, 710–714 (2017).
44. Z. Ma *et al.*, A *Phytophthora sojae* glycoside hydrolase 12 protein is a major virulence factor during soybean infection and is recognized as a PAMP. *Plant Cell* **27**, 2057–2072 (2015).
45. L. Van Valen, "19. A New Evolutionary Law (1973)" in *Foundations of Macroecology: Classic Papers with Commentaries*, F. A. Smith, J. L. Gittleman, J. H. Brown, Eds. (University of Chicago Press, Chicago, 2014), pp. 284–314.
46. G. Bell, *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Routledge, 2019).
47. B. T. Grenfell, A. P. Dobson, H. K. Moffatt, *Ecology of Infectious Diseases in Natural Populations* (Cambridge University Press, 1995).
48. R. D. Finn *et al.*, HMMER web server: 2015 update. *Nucleic Acids Res.* **43**, W30–W38 (2015).
49. R. D. Finn *et al.*, The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
50. M. Blum *et al.*, The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
51. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
52. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
53. I. Letunic, P. Bork, Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
54. B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
55. J. E. Stajich *et al.*, FungiDB: An integrated functional genomics database for fungi. *Nucleic Acids Res.* **40**, D675–D681 (2012).
56. E. Y. Basenko *et al.*, FungiDB: An integrated bioinformatic resource for fungi and oomycetes. *J. Fungi* **4**, 39 (2018).
57. C. J. Van Den Bergh, A. C. A. P. A. Bekkers, P. De Geus, H. M. Verheij, G. H. De Haas, Secretion of biologically active porcine prophospholipase A2 by *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **170**, 241–246 (1987).
58. G. L. Miller, Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **31**, 426–428 (1959).
59. E. R. Master, Y. Zheng, R. Storms, A. Tsang, J. Powlowski, A xyloglucan-specific family 12 glycosyl hydrolase from *Aspergillus niger*: Recombinant expression, purification and characterization. *Biochem. J.* **411**, 161–170 (2008).
60. L. A. Kelley, M. J. E. Sternberg, Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
61. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
62. M. Varadi *et al.*, Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
63. M. N. Wass, L. A. Kelley, M. J. E. Sternberg, 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Res.* **38**, W469–W473 (2010).
64. C. Lamb, R. A. Dixon, The oxidative burst in plant disease resistance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 251–275 (1997).
65. S. C. Amack, M. S. Antunes, CaMV35S promoter–A plant biology and biotechnology workhorse in the era of synthetic biology. *Curr. Plant Biol.* **24**, 100179 (2020).
66. G. Felix, J. D. Duran, S. Volko, T. Boller, Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265–276 (1999).
67. L. Gómez-Gómez, G. Felix, T. Boller, A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *Plant J.* **18**, 277–284 (1999).
68. K. Kuchitsu, H. Kosaka, T. Shiga, N. Shibuya, EPR evidence for generation of hydroxyl radical triggered by N-acetylchitooligosaccharide elicitor and a protein phosphatase inhibitor in suspension-cultured rice cells. *Protoplasma* **188**, 138–142 (1995).
69. Z. Li *et al.*, A C-terminal proline-rich sequence simultaneously broadens the optimal temperature and ph ranges and improves the catalytic efficiency of glycosyl hydrolase family 10 ruminal xylanases. *Appl. Environ. Microbiol.* **80**, 3426–3432 (2014).
70. T.-N. Wen, J.-L. Chen, S.-H. Lee, N.-S. Yang, L.-F. Shyur, A truncated fibrobacter succinogenes 1,3-1,4-β-d-glucanase with improved enzymatic activity and thermotolerance. *Biochemistry* **44**, 9197–9205 (2005).
71. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
72. N. Galtier, M. Gouy, C. Gautier, SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Bioinformatics* **12**, 543–548 (1996).
73. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
74. S. Guindon, O. Gascuel, A. Simple, Fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
75. M. A. Larkin *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
76. F. Madeira *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
77. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2–A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
78. F. Teufel *et al.*, SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).
79. P. J. Wood, J. Weisz, Detection and assay of (1-4)-(8-D-Glucanase,(1-3)-(8-D-Glucanase, (1-3)(1-4)-(3-D-Glucanase), and xylanase based on complex formation of substrate with congo red. *Cereal Chem.* **641**, 8–15 (1987).
80. A. Signorell *et al.*, DescTools: Tools for descriptive statistics (Version 099 45, R Package, 2022). https://andrisignorell.github.io/DescTools/. Accessed 31 August 2022.
81. Student, The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
82. T. Nakagawa *et al.*, Improved Gateway binary vectors: High-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* **71**, 2095–2100 (2007).
83. R. Höfgen, L. Willmitzer, Storage of competent cells for *Agrobacterium* transformation. *Nucleic Acids Res.* **16**, 9877 (1988).
84. L. M. Shaw *et al.*, DirtyGenes: Testing for significant changes in gene or bacterial population compositions from a small number of samples. *Sci. Rep.* **9**, 2373 (2019).
85. V. Attah, X. Yan, G. Leonard, T. A. Richards, Data from "Duplication and Neofunctionalization of a horizontally transferred xyloglucanase as a facet of the Red Queen coevolutionary dynamic." Figshare. https://dx.doi.org/10.6084/m9.figshare.21524886. Deposited 10 April 2024.