Plant Expansins in Bacteria and Fungi: Evolution by Horizontal Gene Transfer and Independent Domain Fusion

Nikolas Nikolaidis,*[1] Nicole Doran,[1] and Daniel J. Cosgrove[2]

[1]Department of Biological Science and Center for Applied Biotechnology Studies, California State University, Fullerton

[2]Department of Biology, Pennsylvania State University, University Park

***Corresponding author:** E-mail: nnikolaidis@fullerton.edu.

Associate editor: Jianzhi Zhang

Abstract

Horizontal gene transfer (HGT) has been described as a common mechanism of transferring genetic material between prokaryotes, whereas genetic transfers from eukaryotes to prokaryotes have been rarely documented. Here we report a rare case of HGT in which plant expansin genes that code for plant cell-wall loosening proteins were transferred from plants to bacteria, fungi, and amoebozoa. In several cases, the species in which the expansin gene was found is either in intimate association with plants or is a known plant pathogen. Our analyses suggest that at least two independent genetic transfers occurred from plants to bacteria and fungi. These events were followed by multiple HGT events within bacteria and fungi. We have also observed that in bacteria expansin genes have been independently fused to DNA fragments that code for an endoglucanase domain or for a carbohydrate binding module, pointing to functional convergence at the molecular level. Furthermore, the functional similarities between microbial expansins and their plant xenologs suggest that these proteins mediate microbial–plant interactions by altering the plant cell wall and therefore may provide adaptive advantages to these species. The evolution of these nonplant expansins represents a unique case in which bacteria and fungi have found innovative and adaptive ways to interact with and infect plants by acquiring genes from their host. This evolutionary paradigm suggests that despite their low frequency such HGT events may have significantly contributed to the evolution of prokaryotic and eukaryotic species.

Key words: **xenology, convergence, plant pathogens, endoglucanases, grass–pollen allergens.**

Introduction

Horizontal gene transfer (HGT), the nonsexual transmission of genetic material across species, is a common and important mechanism in prokaryotes, which influences their function and evolution (Koonin et al. 2001). By contrast, HGT is considered rare between multicellular eukaryotes. However, HGT has been relatively well documented in plants and, in most cases, involves the transfer of genetic material between a host plant and its plant parasite (Bergthorsson et al. 2003; Won and Renner 2003; Davis and Wurdack 2004; Mower et al. 2004; Davis et al. 2005; Richardson and Palmer 2007; Xi et al. 2013). This phenomenon has been attributed to the direct physical association of the parasite and its host, (Bergthorsson et al. 2003; Won and Renner 2003; Xi et al. 2012) and with a few exceptions it involves mainly mitochondrial genes (Yoshida et al. 2010; Xi et al. 2012). It has been proposed that these events confer selective advantages to the parasites, which by mimicking their host genome increase their fitness (Xi et al. 2012). Although HGTs between plants are now considered an important mechanism of evolution, especially for parasitic plants, nuclear HGTs between plants and prokaryotes or other eukaryotes with the plant species being the donor have rarely been documented.

In this study, we investigated whether EXLX1, a *Bacillus subtilis* protein that is structurally and functionally very similar to plant expansins (Kerff et al. 2008; Georgelis et al. 2011, 2012), was acquired from plants by HGT. Plant expansins are proteins involved in cell enlargement and in a variety of other developmental processes involving cell wall modification (McQueen-Mason et al. 1992; Li et al. 1993; Shcherban et al. 1995; Cosgrove 2005). These wall-loosening proteins are typically 250–275 amino acids long and contain two domains: an N-terminal domain (D1) with distant sequence similarity to the catalytic domain of the family-45 endoglucanases and a C-terminal domain (D2) that is related to a family of grass–pollen allergens of unknown function (Li et al. 2002; Cosgrove 2005; Sampedro and Cosgrove 2005; Yennawar et al. 2006). Despite the similarity with endoglucanases, no enzymatic activity has been found that accounts for the action of expansin on the wall (McQueen-Mason et al. 1992; Cosgrove 2000; Li et al. 2003; Yennawar et al. 2006; Kerff et al. 2008; Georgelis et al. 2012). D2 has been shown to have properties similar to a type-A cellulose binding domain, where binding to the hydrophobic surface of cellulose microfibrils is primarily mediated by aromatic residues on the open binding surface (Georgelis et al. 2012).

The mechanism by which plant expansins function remains enigmatic, but these proteins are believed to disrupt noncovalent binding of wall polysaccharides to one another (McQueen-Mason et al. 1992; Cosgrove 2000; Li et al. 2003). Toward identifying the molecular mechanism by which expansins function, we solved the crystal structure of a

maize expansin, EXPB1 (Yennawar et al. 2006), which consists of two domains (D1 and D2) closely packed and aligned so as to form a long, shallow groove with a potential to bind a glycan backbone. We proposed that EXPB1 targets arabinoxylan–cellulose junctions within the wall to promote slippage, stress relaxation, and yielding of the cell wall, without lysis of the cell wall components.

Phylogenetically, plant expansins are classified into four major families: α-expansin (EXPA), β-expansin (EXPB), expansin-like A (EXLA), and expansin-like B (EXLB) (Sampedro and Cosgrove 2005; Sampedro et al. 2006). Although it is known that the EXPA and EXPB families preceded the divergence of vascular plants and mosses and the EXLA and EXLB families preceded the divergence of seed plants (Sampedro and Cosgrove 2005; Sampedro et al. 2006; Carey et al. 2013), the evolutionary origin of the expansin molecule, that is, a protein with homology to both expansin domains (D1 and D2), remains elusive. In other words, it is currently unknown when the fusion of the two domains (D1 and D2) occurred. The main reason for this uncertainty is the fact that only a few nonplant species have proteins with similarity to full-length expansins (both domains) (Li et al. 2002; Sampedro and Cosgrove 2005). With the exception of proteins found in the green alga, *Micrasterias denticulata* (Vannerum et al. 2011), homology between plant expansins and expansin-like proteins of other microbes has not been established, because primary sequence similarity searches produce alignments with very low scores.

Further clues into the evolutionary origin of the nonplant proteins that align with the full-length plant expansins comes from the activity and crystal structure of one of these nonplant proteins, the EXLX1protein of *B. subtilis* (Kerff et al. 2008; Georgelis et al. 2012). Despite low sequence similarity, EXLX1 is remarkably similar in structure to EXPB1, consisting of two tightly packed domains (D1, D2) that form an open surface suitable for binding to a nearly flat polysaccharide surface, such as cellulose. In complex with cellohexaose or related cellulose-like oligosaccharides, EXLX1 forms a novel sandwich-like structure wherein the oligosaccharide is confined between the aromatic residues of the D2 domains of two proteins in the crystallographic unit. EXLX1 binds to plant cell walls via the D2 domain, promotes plant cell wall extension, and lacks lytic activity against a variety of plant cell wall polysaccharides—functional characteristics similar to those of plant EXPBs (Kerff et al. 2008; Georgelis et al. 2011, 2012). Moreover, deletion of the gene encoding EXLX1 greatly reduces the ability of the bacterium to colonize maize roots, supporting a physiological role of EXLX1 in plant–bacterial interactions (Kerff et al. 2008). More recently, other microbial proteins were found to possess similar functional characteristics to both EXLX1 and plant expansins (Georgelis et al. 2013).

These functional and structural analogies between plant expansins and the *B. subtilis* EXLX1 led us to hypothesize that EXLX1 and similar microbial proteins are homologs. Furthermore, taking into account the presence of these expansin-like genes in plant pathogens and several reports that described multiple HGTs between hosts and their pathogens, especially between plants and their parasitic plants (Davis and Wurdack 2004; Mower et al. 2004; Davis et al. 2005; Xi et al. 2012), we hypothesized that EXLX1 has been acquired by bacteria and other microbes through HGT from plants. To test these hypotheses, we investigated the origin and evolution of the EXLX1 protein by sequence-based phylogenetic analysis.

Results

Homologs of EXLX1 in Nonplant Species

To test whether proteins homologous to EXLX1 exist in other species, we searched the nonredundant (nr) Database of National Center for Biotechnology Information (NCBI) with protein Basic Local Alignment Search Tool (BLASTp) and tBLASTn using the EXLX1 sequence as query. The criteria of these searches were at least 80% coverage of the query sequence and E-values lower than $1e^{-04}$. Using these criteria, we identified several sequences from bacteria, fungi, and unicellular eukaryotes (amoebozoa) that align to both domains of EXLX1 (fig. 1 and supplementary table S1, Supplementary Material online). These alignments had low E values, ranging from $7e^{-170}$ to $1e^{-04}$ and high scores (481–65.5), suggesting that these proteins are homologous to EXLX1 (supplementary table S1, Supplementary Material online). To further test this inference, we used domain and fold recognition programs. These analyses showed that all the identified proteins contain both D1 and D2 domains like EXLX1 and EXPB1, and the predicted folds are very similar to both EXLX1 and EXPB1 (fig. 2 and data not shown). Additionally, analysis of the conservation levels suggests that all proteins code for a conserved surface similar to both EXLX1 and EXPB1 (fig. 2). Furthermore, specific amino acid sites are conserved between the microbial sequences and EXLX1 and EXPB1 (supplementary fig. S1, Supplementary Material online). In combination, these data support sequence homology between EXLX1 and the identified proteins in bacteria and other nonplant eukaryotes.

Three major observations can be deduced from the phylogenetic analysis of EXLX1 homologs. First, the distribution of the EXLX1 homologs is sporadic (figs. 1*a* and 3). Second, we observed that in some cases the EXLX1 homologous proteins contain additional domains (fig. 1*b*). In all cases, these domains are either cellulase GH5 or carbohydrate-binding modules (CBMs). Third, many of the EXLX1 homologs exist in species that are plant pathogens (40% of species) (fig. 1*a*). Most of the remaining proteins exist in species that either live in soil having a direct relationship with plants (e.g., *Bacillus*, *Frankia*) or produce cellulose during their life cycle (e.g., *Dictyostelium*).

Species Distribution of EXLX1 Homologs

The EXLX1 homologs are found only in a few bacterial species (3% of the sequenced bacterial genomes; supplementary table S2, Supplementary Material online), a few fungal species (5% of the sequenced fungal genomes; supplementary table S3, Supplementary Material online), and in amoebozoa from other eukaryotes. To ensure that all homologs have been identified, we used tBLASTn and searched the nr

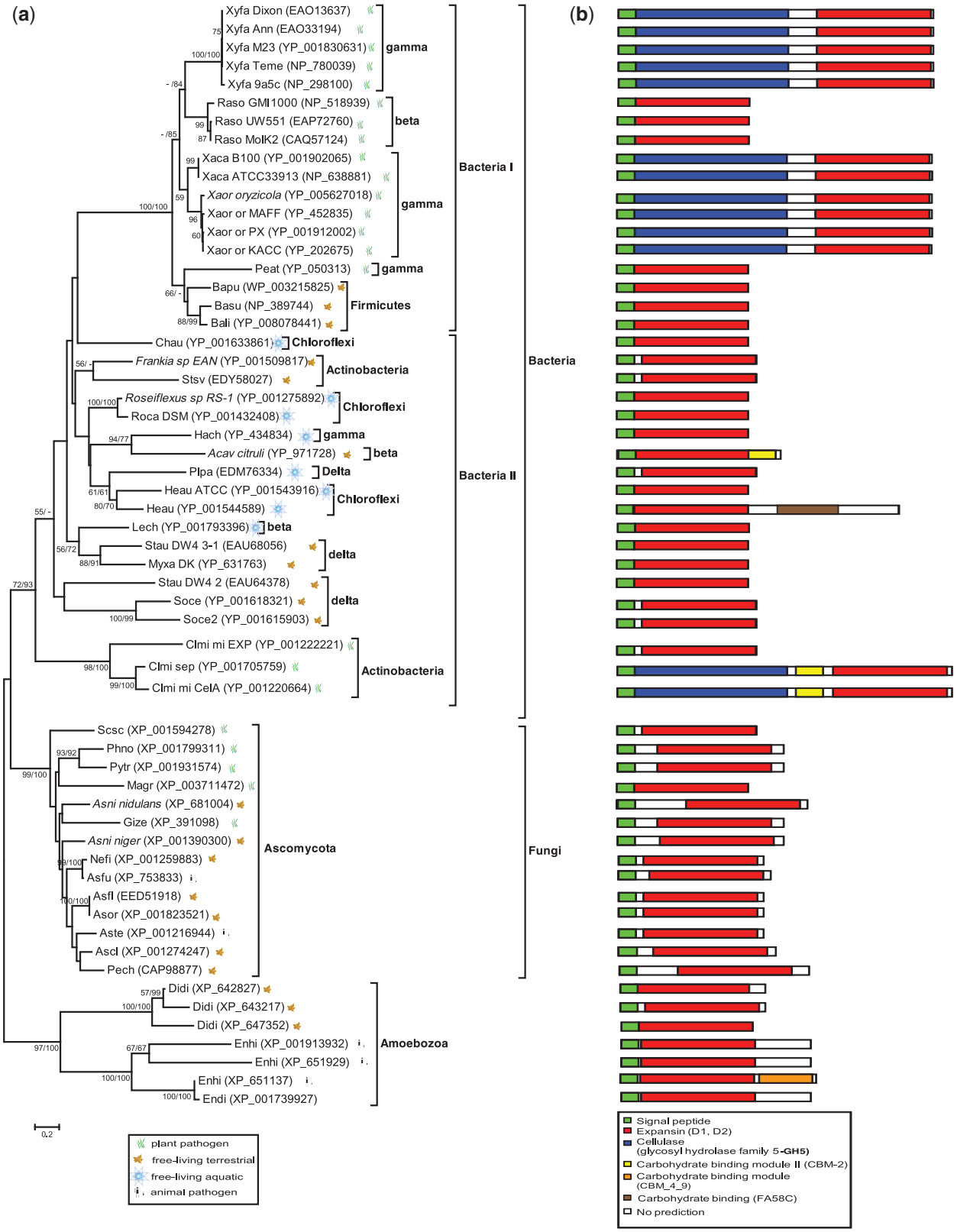

FIG. 1. The *Bacillus subtilis* EXLX1protein has many homologs in bacteria, fungi, and amoebozoa. (*a*) Phylogenetic relationships of representative *B. subtilis* EXLX1 homologs. Two different phylogenetic methods (NJ and ML) were used with gamma-distributed distances from the WAG substitution model with $\alpha = 1.72$. Alignment gaps were excluded and the total number of sites used to construct the trees was 176. The numbers at the nodes are bootstrap values (NJ/ML). The biology of each species is shown with different symbols next to the species name. Species names abbreviations are given in supplementary table S1, Supplementary Material online. Only sequences producing BLAST hits with E-values lower than 10^{-4} and query coverage higher than 80% were used for the construction of these trees (*b*). Many EXLX1homologs contain additional domains. The domain organization of the EXLX1 homologs was identified using the Conserved Domains Database (CDD) database from NCBI coupled to fold recognition analysis. We define as expansin the domain that contains both D1 and D2 domains according to the EXLX1structure (Kerff et al. 2008; Georgelis et al. 2013).

taxon-specific database using as queries representative sequences from the different phylogenetic groups of figure 1. These searches did not reveal additional sequences that contain both domains (more than 80% query coverage), with the exception of plant expansins. In particular, we found that when we used as queries sequences from bacteria group II, fungi, or amoebozoa (fig. 1*a*), we retrieved plant expansin proteins with significant alignment scores and low E-values (supplementary table S1, Supplementary Material online). We also observed that nonplant expansins are the best reciprocal BLAST hits of plant expansins. These analyses together with the remarkable similarities between the EXPB1 and EXLX1 structures (fig. 2 and data not shown; see also Kerff et al. 2008 and Georgelis et al. 2012) suggest homology between plant and nonplant expansins.

We then relaxed the cutoff values allowing less than 80% coverage of the query sequence and collected sequences that produced alignments with very low scores (supplementary table S1, Supplementary Material online). By these means, we found a number of proteins that showed similarities to domain 1 (D1) alone and a few proteins with similarities to domain 2 (D2) only (supplementary table S1, Supplementary Material online). We further analyzed these proteins using phylogenetic and fold recognition methods, and we found that the sequences aligned to D2 all belong to the swollenin family of proteins from fungi (supplementary figs. S2 and S3, Supplementary Material online). Our analysis also suggested that these proteins encode an N-terminal domain that folds similarly to D1 of expansins, but it contains many conserved insertions found neither to expansins nor to other Barwin-like

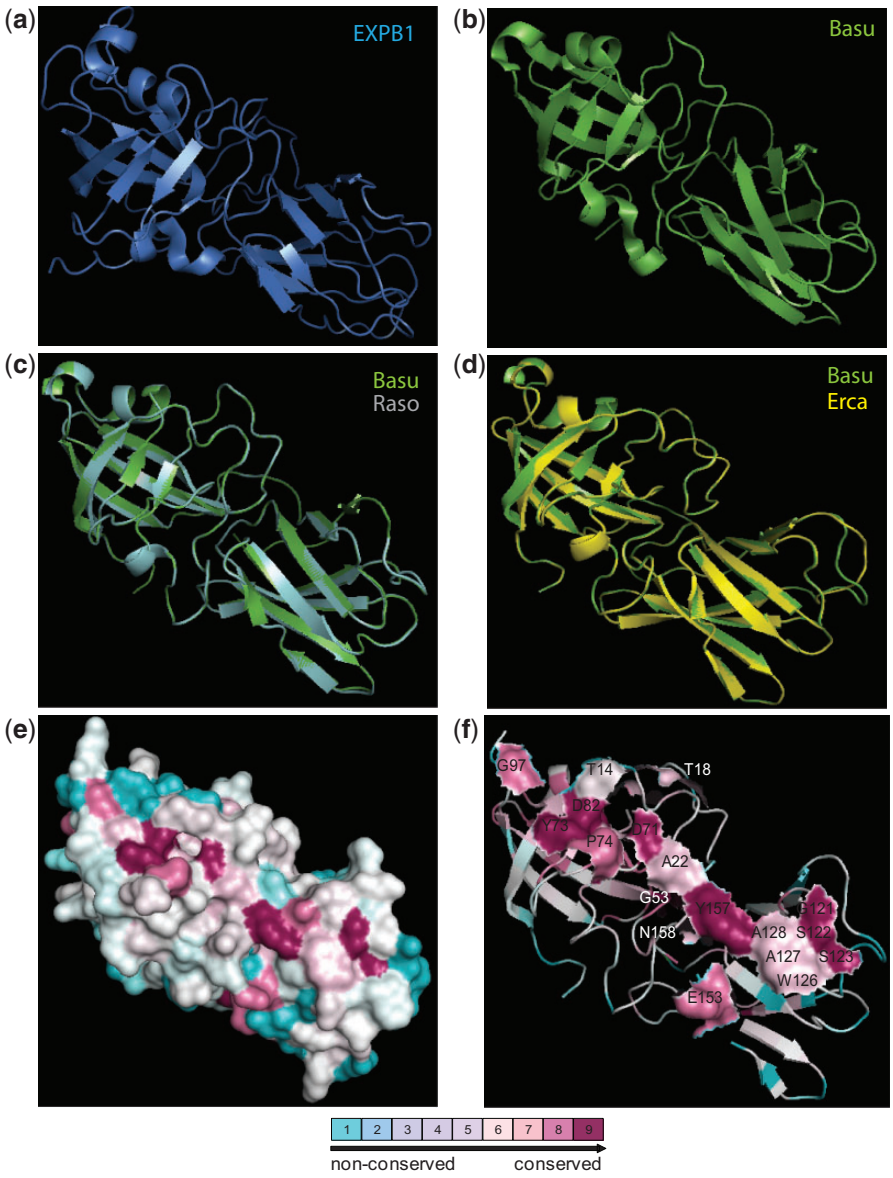

Fig. 2. The EXLX1 homologs are predicted to contain two domains, fold similarly to the *Zea mays* EXPB1 (*a*) and the *B. subtilis* EXLX1 (*b*), and contain a conserved long hydrophobic surface. (*c*, *d*) Structural alignments of the three-dimensional models of the EXLX1homologs from *Ralstonia* and *Erwinia* with the EXLX1structure. Surface (*e*) and ribbon (*f*) representations of the EXLX1structure are colored according to conservation in 70 EXP domain sequences from bacteria and fungi (blue→red with increasing conservation).

endonuclease domains. These data suggest that swollenins are homologous to expansins. Additionally, we observed a few more sequences present in animals, e.g., *Crassostrea gigas* and *Nematostella vectensis* (supplementary table S1, Supplementary Material online). Although these sequences seem to contain both domains, their relationships to each other and to expansins are not clear and unresolved (data not shown).

Analysis of the proteins that aligned only to D1 of EXLX1 revealed that these proteins either do not encode a C-terminal domain or this domain is not homologous to the D2 domain of EXLX1 (data not shown). The evolutionary relationships between D1 only containing proteins and expansins are ambiguous because most of them clustered all together as outgroups to the expansin clade (figs. 4 and 5).

Evolution of EXLX1 Homologs

The analyses of the tertiary and the primary sequences strongly support the conclusion that EXLX1 has many homologs in phylogenetically diverged species. Our phylogenetic analysis reveals that the tree derived from the EXLX1 and its homologs in both bacteria and fungi is different from the species trees (fig. 3 and supplementary fig. S3, Supplementary Material online). A plausible explanation of the observed phylogenetic pattern of the EXLX1 homologs is HGT between different bacterial and fungal species. The alternative explanation of independent fusion of the two

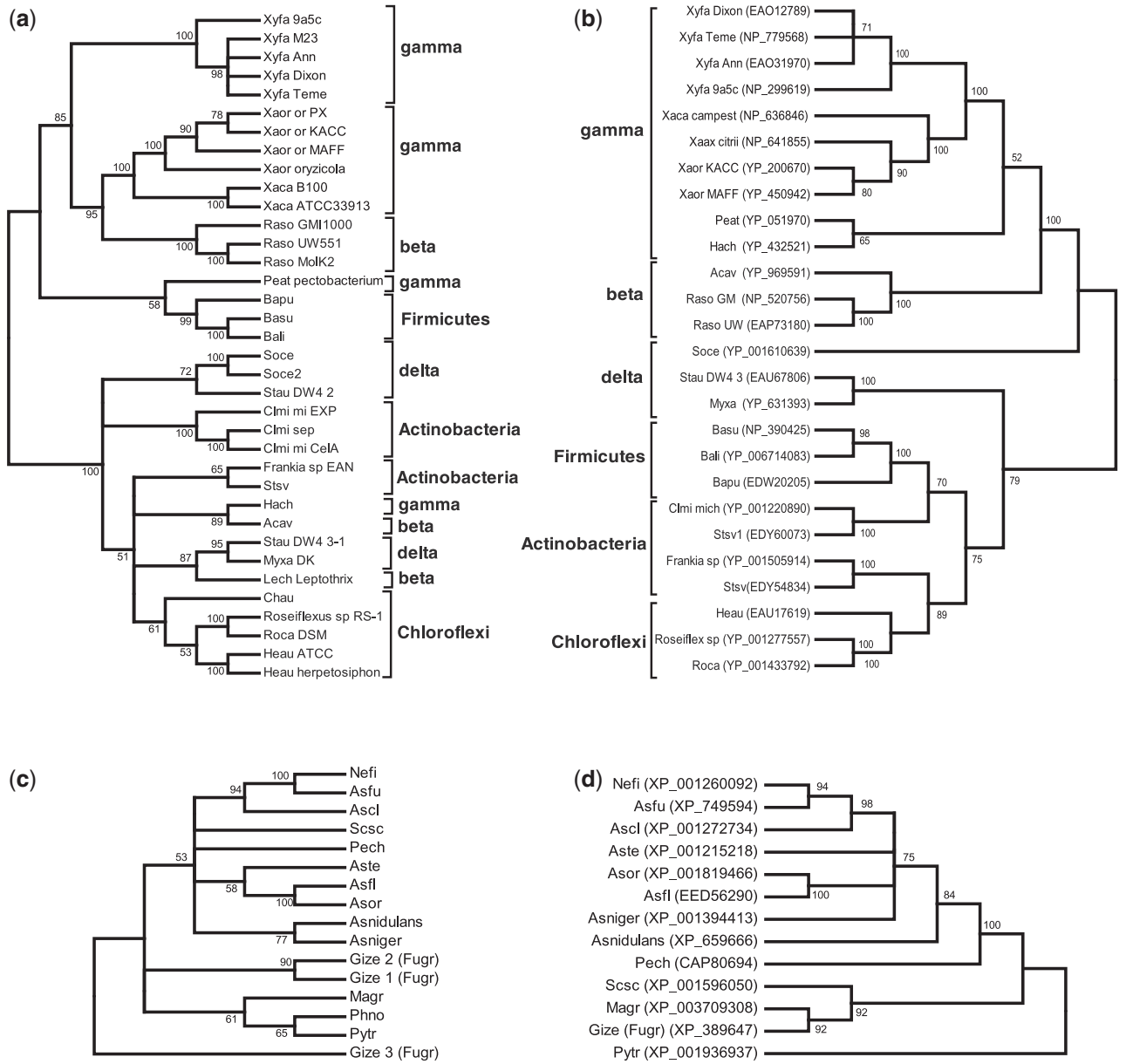

Fig. 3. The phylogenetic relationships of the EXLX1homologs in bacteria and fungi are different from the species trees. (*a*, *b*). Phylogenetic trees of the EXLX1homologs and the DNAK homologs in selected bacterial species. (*c*, *d*). Phylogenetic trees of the EXLX1homologs and the DNAK homologs (ER protein—single copy gene) in selected fungal species. All trees were constructed using *p* distances after complete elimination of alignment gaps and were drawn by the NJ method.

domains (D1 and D2) seems rather implausible because it would require multiple independent events. This is particularly true considering expansin's tightly packed two-domain structure with a conserved open surface spanning both domains, which is part of the definition of canonical expansins (Sampedro and Cosgrove 2005). Therefore, the most plausible scenario that explains the evolution of these proteins is ancient HGTs followed by vertical transmission in some instances (e.g., *Aspergillus*, amoebozoa species). To further explore this notion, we analyzed several parametric characteristics (GC content, GC skew, as well as codon and amino acid usage) of the EXLX1 homologs, because in recent HGT cases, the transferred genes retain some characteristics of the donor species. Our results argue against recent HGT events, because we found no major differences in any of the parametric measures analyzed (data not shown). Therefore, ancient HGT events are more likely (Koonin et al. 2001; Richardson and Palmer 2007; Ni et al. 2012).

One aspect that could be attributed to convergence are the fusions between expansin domains and homologous cellulase GH5 domains, which were found in the genomes of three phylogenetically divergent bacterial species (fig. 1). The comparison of the topologies produced separately by the expansin and the cellulase domains strongly suggests that these fusions occurred independently in these species (fig. 1 and supplementary fig. S4, Supplementary Material online). The alternative explanations of vertical inheritance or horizontal transfer of the fused proteins are not supported by our data, because the former explanation requires massive losses of genes in all intermediate species and the latter is not supported by the topology of the cellulase domains. The explanation of independent origin of fused proteins with homologous domains also applies for the fusion of the expansin domain (both D1 and D2) to different types of CBMs in fungi and amoebozoa (fig. 1 and supplementary figs. S5 and S6, Supplementary Material online). Fusions between an expansin domain and CBMs are also observed in *Entamoeba histolytica* (fig. 5), which contains one protein that encodes a CBM_4_9 domain and *Herpetosiphon aurantiacus*, which encodes a protein with a FA58C domain. The latter domain is present in eukaryotes and is assumed to have been transferred horizontally to eubacterial genomes (Baumgartner et al. 1998). Although the function of these modular proteins is largely unknown, some reports have suggested that expansin domains may act synergistically with cellulase domains to augment the hydrolytic activities of the latter (Kim et al. 2009), but this suggestion is refuted by other results showing no synergistic action (Kerff et al. 2008; Georgelis et al. 2011, 2012). In a recent study, we found that the expansin domain of a modular GH5 endoglucanase of *Xanthomonas campestris* (Xaca) does not significantly affect the endoglucanase activity under the conditions tested (Georgelis et al. 2013). Given these results, we speculate that the expansin domain could act as an anchor to localize the endoglucanase domain to selective regions in the cell wall. This concept is consistent with recent nuclear magnetic resonance (NMR) results showing that *B. subtilis* EXLX1 binds to a form of cellulose distinctive from the bulk of the cellulose in the cell wall (Wang et al. 2013).

Origin of the Expansin Domain

Our data showed that the *B. subtilis* EXLX1 protein has many homologs in phylogenetically diverged bacteria, fungi, and amoebozoa. Furthermore, our results suggested that these proteins (or domains within modular proteins) are homologous to plant expansins. We next wondered when and in which organismal lineage the expansin domain appeared. The origin of the expansin domain can be explained by two alternative scenarios. First, the expansin domain may have originated independently in plants, bacteria, or fungi and was then transmitted by HGT. This hypothesis is not supported by our data, which provide evidence for homology rather than analogy between the expansin domain of bacteria, fungi, amoebozoa, and plants. Furthermore, if we assume multiple independent events, then we should be able to identify the two ancestral modules that produced the expansin domain in multiple lineages. We found that many bacterial and eukaryotic genomes encode domains similar to the expansin D1 (structurally similar to Barwin-like GH45 endoglucanases and related proteins), but a domain similar to the expansin D2 could be identified only in expansins (figs. 4 and 5 and supplementary fig. S2, Supplementary Material online). Also, domain D1 of plant expansins is evolutionary and more related to bacterial and fungal expansins than to other Barwin-like endoglucanases (fig. 4). These results do not support the hypothesis of independent fusion. Furthermore, it seems highly improbable that a compact protein consisting of two tightly packed domains with identical folds and forming a conserved surface spanning both domains would have been invented independently multiple times in evolution. Collectively, our data strongly suggest that the expansin domain arose only once through evolution.

The second scenario assumes that the expansin domain originated in bacteria or fungi or in the common ancestor of plants and algae and then was vertically or horizontally transmitted to other species. The vertical transmission hypothesis seems rather implausible, because it would require massive losses of genes from multiple prokaryotic and eukaryotic lineages. Therefore, HGT must have played an important role in the distribution of these proteins in many diverse phyla. Although the remarkable similarities between plant and bacterial expansin structures (fig. 2 and data not shown; Kerff et al. 2008; Georgelis et al. 2012) as well as the tree topologies strongly support the HGT model of evolution, they do not unequivocally determine which organism is the donor (figs. 1 and 4). In particular, although, the topology produced by the D1 domain sequence favors the plant origin scenario (fig. 4), it does not provide conclusive evidence for the orientation, time, and number of these events. The fact that expansins exist in all plants (universal representation; Sampedro and Cosgrove 2005) while they are present only in a few bacterial and fungal species (sporadic representation) strongly suggests that the expansin domain originated in plants and was

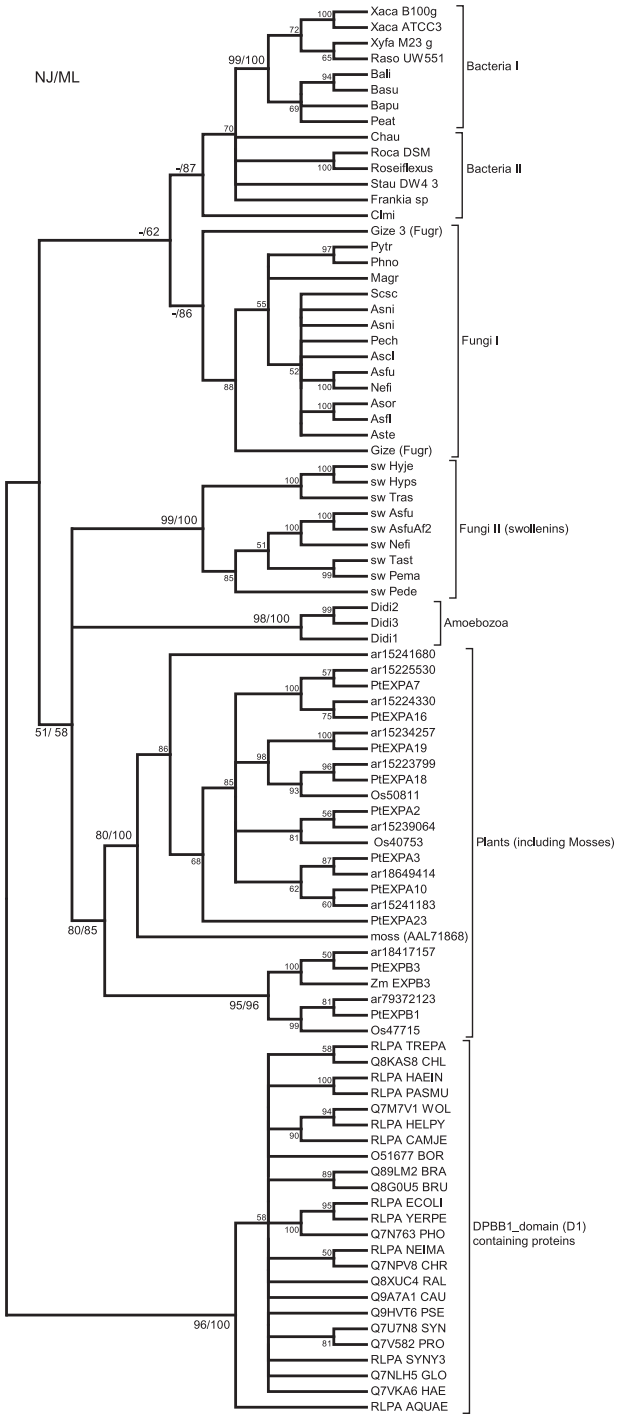

Fig. 4. The N-terminus domain (D1) of plant expansins is more related to bacterial and fungal expansins than is to other Barwin-like endoglucanases. Two different phylogenetic methods (NJ and ML) were used with gamma-distributed distances from the WAG substitution model with $\alpha = 1.063$. Alignment gaps were excluded, and the total number of sites used to construct the trees was 76. The numbers at the nodes are bootstrap values (NJ/ML). Species abbreviations: Ar, *Arabidopsis thaliana*; Pt, *Populus trichocarpa*; Os, *Oryza sativa*; moss, *Physcomitrella patens*; abbreviations for the other species are given in supplementary table S1, Supplementary Material online. The DPBB1-domain sequences corresponding to D1 domain of expansins were downloaded from the PFAM database (ID: pfam03330).

transmitted to other groups by HGT. The HGT scenario is also indirectly supported by the presence of an expansin domain in microbes that are either colonizing plants or are plant pathogens.

The reductio ad absurdum (proof by contradiction) arguments presented above—favoring a unique origin of the expansin protein in the common ancestor of plants and its subsequent horizontal transmission to nonplant species—are solely based on sequence analysis. The functional activities of these proteins provide additional evidence to support this scenario. Our reasoning was that if plant expansins have been transferred to bacteria, then we would be able to detect unique functional similarities between the xenologs. Indeed, in our recent paper, we demonstrate that expansins from three plant-pathogenic bacteria and one fungus cause extension of cell walls in vitro and weaken filter paper networks and possess no lytic activity (Georgelis et al. 2013). These functional similarities of the microbial expansins with plant expansins, including the lack of catalytic (enzymatic) activity, argues against an independent fusion of D1 and D2, because such a scenario would require the independent loss and gain of the same amino acid residues. Thus, these considerations strongly suggest that microbial and plant expansins are xenologs. Together with the sequencing analyses, the functional data provide additional support to the scenario in which the expansin fold was invented only once through evolution in the common ancestor of plants and was then transferred in nonplant species by HGT.

Discussion

Our results show that the *B. subtilis* protein EXLX1 is the prototypic member of a gene family with homologs in diverse bacterial and fungal species. Furthermore, the sequence and structural analyses together with the functional data of microbial proteins suggest that the EXLX1 homologs in bacteria, fungi, and amoebozoa are homologous to plant expansins. The most plausible explanation of our results is that the expansin domain arose once in evolution at the common ancestor of plants and was horizontally transmitted to other nonplant species. This scenario is further supported by the presence of the emerging plant pathogen *Streptomyces acidiscabies* of a sequence that clusters with plant expansins and is much more similar to plant expansins than any other proteins (supplementary fig. S7, Supplementary Material online). It is possible that this gene is an example of an HGT event so recent that relatively little sequence divergence has occurred. This species also possesses an EXLX1 homolog, that is, with a sequence highly divergent from plant expansins.

The alternative explanation of vertical inheritance of the expansin domain from bacteria to other species would require numerous gene losses in both prokaryotes and eukaryotes. Also, the convergent evolution of the expansin domain is not supported by our findings, because in the case of expansins, evolution would have to independently fuse the same two domains multiple times to generate the same unique compact structure with a conserved surface spanning both of them. Furthermore, the activity of microbial

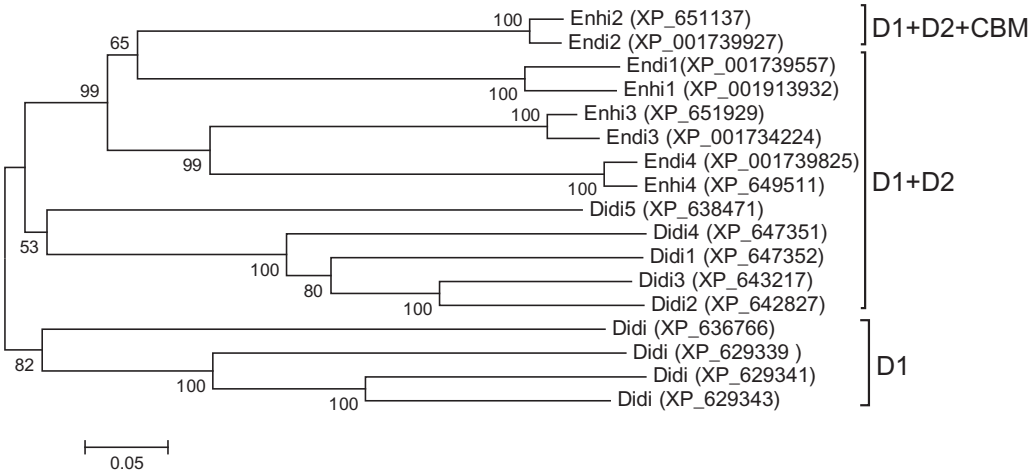

Fig. 5. Phylogenetic relationships of the amoebozoan expansins and related sequences encoding only the D1 domain of expansins. The tree was constructed using *p* distances after complete elimination of alignment gaps and was drawn by the NJ method. Note the fusion of the EXP domain with a CBM domain only in a single protein. Species abbreviations: Enhi, *Entamoeba histolytica*; Endi, *Entamoeba dispar*; Didi, *Dictyostelium discoideum*.

expansins (plant–cell wall extension, filter paper weakening, cellulose binding, yet no lytic activity) make the independent origin hypothesis improbable, since in addition to the domain fusion it would require the loss or gain of the same amino acids in several different lineages. Indirect support to the improbable nature of such events comes from a wall-loosening protein in the nematode *Globodera rostochiensis* (Smant et al. 1998). This molecule most probably represents an independent fusion between a CBM domain and a Barwin-like endoglucanase domain (similar to EXPB1 D1 domain). The result was a protein with reversed orientation (N-terminal CBM and a C-terminal EXPB1-like D1) when compared to bacterial and plant expansins and did not result in an expansin fold (data not shown), yet is reported to have expansin-like activity. Collectively, our study describes a rare case of HGT in which eukaryotes are the donors. Such HGT cases are less well documented bacterial to bacterial or bacterial to eykaryotes HGT cases (Koonin et al. 2001; Dunning Hotopp et al. 2007; Keeling and Palmer 2008). Furthermore, the current study documents the apparently rare case of a nuclear gene in plants being transferred to prokaryotes, whereas most documented cases of HGT between plants involve genes of mitochondrial origin (Richardson and Palmer 2007; Richards et al. 2009; Yoshida et al. 2010; Xi et al. 2012).

Difficult as it might be to prove beyond reasonable doubt that HGT has occurred during the evolution of the expansins in plants, bacteria, and fungi, it is even harder to unequivocally determine which organism was the donor and which one was the recipient, as well as when and how such events occurred. To begin with, the available collection of genome sequences, in particular plant ones, is but a tiny sampling of the genome universe, because of which we cannot reasonably hope to identify the true source of any gene present in a given genome. At best, it might be possible to propose a credible hypothesis as to the donor lineage. The logic used to formulate such hypotheses assumes that if horizontal transfer has indeed occurred, the taxon with the most diverse representation of the given family is the most likely source. Therefore, lacking a better alternative, we assume that the conception of the expansin fold occurred in the common ancestor of plants and the expansin molecule was transferred horizontally from plants to other nonplant species. The difficulty to determine precisely the donor species makes this particular instance of HGT different from other cases where the donor and recipient species are clearly well defined. The latter is true for several bacterial and plant HGT events, especially between parasitic species and their hosts (Koonin et al. 2001; Bergthorsson et al. 2003; Won and Renner 2003; Mower et al. 2004; Davis et al. 2005; Richardson and Palmer 2007; Rogers et al. 2007; Yoshida et al. 2010; Xi et al. 2012, 2013).

Although HGT can reasonably explain our data, the mechanism by which this transfer occurred remains a mystery. Taking into account that several recipient species are plant pathogens, we suspect that the intimate association between them and their hosts may have facilitated these genetic transfers. Such mechanisms have been suggested to explain the host-to-parasite and parasite-to-host HGT events in several plant species (Bergthorsson et al. 2003; Won and Renner 2003; Mower et al. 2004; Davis et al. 2005; Richardson and Palmer 2007; Rogers et al. 2007; Yoshida et al. 2010; Xi et al. 2012, 2013). At the molecular level, we can only speculate that mechanisms similar to the ones described for other cases of HGT (Richardson and Palmer 2007; Keeling and Palmer 2008; Nedelcu et al. 2008; Ni et al. 2012) could have functioned in the case of expansins. The presence of an expansin gene in a plasmid (i.e., *Clavibacter*; Laine et al. 2000) points to a plasmid mediated transfer as a potential mechanism.

Furthermore, it remains unclear whether plant expansin gene transfers occurred once or several times. Taking into account the topologies shown in figures 1 and 4, the presence or absence of conserved cysteines in bacterial expansins (supplementary fig. S1, Supplementary Material online), and the presence of multiple expansin genes in some species (e.g., *Stigmatella*, *Clavibacter*, *Giberella*), we suggest that there were at least two HGT events from plants to nonplants. The one gave rise to present day amoebozoa expansins and fungal swollenins and the other gave rise to the bacterial and fungal expansins. These two major events were followed by

multiple secondary HGT events within fungi and bacteria. The *S. acidiscabies* expansin is the result of a different, apparently more recent, HGT event, most probably from a member of the EXPA plant subfamily. The lack of parametric similarities, for example, GC content or codon usage, between plant and nonplant expansins, even for the *S. acidiscabies* gene, as well as between the bacterial and fungal molecules, suggests that these events must have occurred very early during the evolution of these species.

In addition to the abundant HGT between eukaryotes and prokaryotes and between prokaryotes, the evolution of EXLX1 and its nonplant homologs is characterized by several independent fusions between the expansin domain and cellulase GH5 or CBM domains. Although domain fusions are not uncommon among bacteria (Enright et al. 1999), they are less common among eukaryotes and certainly have not been described in plant expansins. Moreover, these events are unique because some of them represent cases of convergent evolution, where a homologous GH5 domain fused independently to an expansin domain. In addition to the convergent evolution that resulted in modular proteins, the fact that in all cases these tinkering events (fusions) include a CBM or a GH5 domain is intriguing, given the functional implications of such events.

The distribution of the EXLX1 homologs in many species that are either plant pathogens or interact with plants (fig. 1) raises the question whether there is an adaptive advantage for these species to make use of an expansin protein. The answer to this question can only come from experiments that will identify the physiological role of these proteins in nonplant species. Gene deletion studies of the *B. subtilis* EXLX1 (Kerff et al. 2008) and the swollenin of the saprophytic fungus *Trichoderma reesei* (Brotman et al. 2008) demonstrated that these proteins promote colonization of plant roots. These findings strongly suggest that despite the high degree of sequence divergence, the two proteins have retained the primordial function of attaching to and perhaps modifying plant cell walls. We speculate that in other species these proteins may function in a similar way to assist the interaction with plants by facilitating the adhesion of bacterial or fungal cells to plant cells. This speculation is indirectly supported by the fact that recombinant bacterial expansins bind both plant and bacterial cell walls (Kerff et al. 2008; Lee et al. 2010; Georgelis et al. 2011, 2012, 2013) and that the *Clavibacter* expansin domain plays an important role in virulence (Jahr et al. 2000). In nonpathogenic species or species that are not known to interact with plants, these proteins may have been coopted to accommodate novel cellular processes, as in *Aspegillus* and *Dictyostellium* (Bouzarelou et al. 2008; Ogasawara et al. 2009). The absence of expansins from quite a few plant-interacting bacteria for which genomic sequences are available suggests that these proteins are not essential for all bacteria–plant interactions. Therefore, we believe that the expansin proteins in plant-interacting microbes may have contributed new or alternative mechanisms of interaction (i.e., infection, colonization) between microbes and plants. The latter supposition, which implies an adaptive advantage for the plant-interacting or plant-infecting organisms containing an expansin gene, is in accordance with a model proposed to explain the selective advantage of parasitic plants acquiring genes from their hosts (Xi et al. 2012).

The evolutionary patterns observed in EXLX1 and its plant and nonplant homologs provide a unique case of evolution between pathogen and host and between different species to interact with each other. Similar to parasitic plants, which have acquired genes from their hosts to use to their advantage (Xi et al. 2012), the acquisition of expansins by bacteria and fungi may have provided them new venues to interact with and infect plants. Our data together with other reports showing the contribution of HGT to the emergence of new diseases and pathogenicity (Nedelcu et al. 2008; Kado 2009; Keeling 2009) suggest that although these events are rare it may have played a significant role in the evolution of both prokaryotes and eukaryotes.

Materials and Methods

Data Collection

The BLASTp and tBLASTn programs (Johnson et al. 2008) were used for the retrieval of gene sequences similar to the EXP domain from the nonredundant (nr; May 2013) database of NCBI as well as from species specific databases (see also supplementary table S1, Supplementary Material Online). The parameters used were E value threshold: 10^{-04}–10; word size: 3; amino acid substitution matrix: BLOSUM45; gap opening cost: 15; gap extension cost: 2; and filter for low-complexity regions: ON. The crystal structures of EXLX1 and EXPB1 were used as queries in the Protein Data Bank (PDB) database (http://www.rcsb.org, last accessed November 5, 2013) in the search for three-dimensional (3D) structures similar to the EXP domain.

Multiple Sequence Alignment

The collected sequences were aligned using the Multiple Alignment using Fast Fourier Transform (MAFFT) program (Katoh and Standley 2013). The parameters used were L-INS-i (iterative refinement method that incorporates local and pairwise alignment information); amino acid substitution matrix: BLOSUM62; gap-opening penalty: 1.53; and offset value: 0.00. The alignments were inspected and manually edited in the sequence editor BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html, last accessed November 5, 2013).

Phylogenetic Analysis

The model of protein evolution that best fits the expansin domain multiple sequence alignment was selected by using the ProtTest program (Abascal et al. 2005). The ProtTest program was also used for estimation of the proportion of invariable sites and the alpha parameter of the gamma-distributed substitution rates. Molecular evolutionary analyses were conducted using the MEGA (version 5.2) (Tamura et al. 2011) and the PHYML (Guindon et al. 2009) programs. The accuracy of the reconstructed trees was examined by the bootstrap test with 1,000 replications in the neighbor-joining

(NJ) method and 100 replications in the maximum likelihood (ML) method.

3D Protein Analysis

Homology modeling was carried out using the Swiss-Model server (Bordoli et al. 2009), using as templates the experimentally resolved structures of EXLX1 (PDB IDs: PDB ID: 3D30, 2BH0, 4FG4) and EXPB1 (PDB ID: 2HCZ). Pairwise structural alignments and structural superimposition were performed using the DALI server (Holm and Park 2000). Identification of functional protein regions was performed using the ConSurf Web server (Ashkenazy et al. 2010). In the ConSurf program, we used the ML method for calculating the amino acid conservation scores. The multiple sequence alignment and the tree files were provided as input attributes. Models and figures were drawn using Pymol (DeLano Scientific).

Supplementary Material

Supplementary figures S1–S7 and tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by United States Department of Energy Grant (DE-FG02-84ER13179 to D.J.C.) from the Office of Basic Energy Sciences, by start-up funds from California State University, Fullerton, and a California State Mini Grant to N.N., and by a grant from the Faculty Development Center at CSUF to N.D. The authors also thank the two reviewers for their valuable comments and suggestions.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38: W529–W533.

Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. 1998. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci.* 7: 1626–1631.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197–201.

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. 2009. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc.* 4:1–13.

Bouzarelou D, Billini M, Roumelioti K, Sophianopoulou V. 2008. EglD, a putative endoglucanase, with an expansin like domain is localized in the conidial cell wall of *Aspergillus nidulans*. *Fungal Genet Biol.* 45: 839–850.

Brotman Y, Briff E, Viterbo A, Chet I. 2008. Role of swollenin, an expansin-like protein from *Trichoderma*, in plant root colonization. *Plant Physiol.* 147:779–789.

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 Biol.* 13:4.

Cosgrove DJ. 2000. Loosening of plant cell walls by expansins. *Nature* 407:321–326.

Cosgrove DJ. 2005. Growth of the plant cell wall. *Nat Rev Mol Cell Biol.* 6: 850–861.

Davis CC, Anderson WR, Wurdack KJ. 2005. Gene transfer from a parasitic flowering plant to a fern. *Proc Biol Sci.* 272:2237–2242.

Davis CC, Wurdack KJ. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* 305: 676–678.

Dunning Hotopp JC, Clark ME, Oliveira DC, et al. (20 co-authors). 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317:1753–1756.

Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90.

Georgelis N, Nikolaidis N, Cosgrove DJ. Forthcoming 2013. Biochemical analysis of expansin-like proteins from microbes. *Carbohydr Polym.*

Georgelis N, Tabuchi A, Nikolaidis N, Cosgrove DJ. 2011. Structure-function analysis of the bacterial expansin EXLX1. *J Biol Chem.* 286: 16814–16823.

Georgelis N, Yennawar NH, Cosgrove DJ. 2012. Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc Natl Acad Sci U S A.* 109:14830–14835.

Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol.* 537:113–137.

Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567.

Jahr H, Dreier J, Meletzus D, Bahro R, Eichenlaub R. 2000. The endo-beta-1,4-glucanase CelA of *Clavibacter michiganensis* subsp. *michiganensis* is a pathogenicity determinant required for induction of bacterial wilt of tomato. *Mol Plant Microbe Interact.* 13:703–714.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36: W5–W9.

Kado CI. 2009. Horizontal gene transfer: sustaining pathogenicity and optimizing host-pathogen interactions. *Mol Plant Pathol.* 10: 143–150.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.

Keeling PJ. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Curr Opin Genet Dev.* 19:613–619.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 9:605–618.

Kerff F, Amoroso A, Herman R, et al. (11 co-authors). 2008. Crystal structure and activity of *Bacillus subtilis* YoaJ (EXLX1), a bacterial expansin that promotes root colonization. *Proc Natl Acad Sci U S A.* 105:16876–16881.

Kim ES, Lee HJ, Bang WG, Choi IG, Kim KH. 2009. Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. *Biotechnol Bioeng.* 102:1342–1353.

Koonin EV, Makarova KS, Aravind L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol.* 55:709–742.

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. *Physiol Mol Plant Pathol.* 57: 221–233.

Lee HJ, Lee S, Ko HJ, Kim KH, Choi IG. 2010. An expansin-like protein from *Hahella chejuensis* binds cellulose and enhances cellulase activity. *Mol Cells.* 29:379–385.

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 Physiol.* 128: 854–864.

Li Y, Jones L, McQueen-Mason S. 2003. Expansins and cell growth. *Curr Opin Plant Biol.* 6:603–610.

Li Z-C, Durachko DM, Cosgrove DJ. 1993. An oat coleoptile wall protein that induces wall extension in vitro and that is antigenically related to a similar protein from cucumber hypocotyls. *Planta* 191: 349–356.

McQueen-Mason S, Durachko DM, Cosgrove DJ. 1992. Two endogenous proteins that induce cell wall extension in plants. *Plant Cell* 4: 1425–1433.

Mower JP, Stefanovic S, Young GJ, Palmer JD. 2004. Plant genetics: gene transfer from parasitic to host plants. *Nature* 432:165–166.

Nedelcu AM, Miles IH, Fagir AM, Karol K. 2008. Adaptive eukaryote-to-eukaryote lateral gene transfer: stress-related genes of algal origin in the closest unicellular relatives of animals. *J Evol Biol.* 21: 1852–1860.

Ni T, Yue J, Sun G, Zou Y, Wen J, Huang J. 2012. Ancient gene transfer from algae to animals: mechanisms and evolutionary significance. *BMC Evol Biol.* 12:83.

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. *Dev Growth Differ.* 51:109–122.

Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21:1897–1911.

Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. *J Exp Bot.* 58:1–9.

Rogers MB, Patron NJ, Keeling PJ. 2007. Horizontal transfer of a eukaryotic plastid-targeted protein gene to cyanobacteria. *BMC Biol.* 5:26.

Sampedro J, Carey RE, Cosgrove DJ. 2006. Genome histories clarify evolution of the expansin superfamily: new insights from the poplar genome and pine ESTs. *J Plant Res.* 119:11–21.

Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. *Genome Biol.* 6:242.

Shcherban TY, Shi J, Durachko DM, Guiltinan MJ, McQueen-Mason SJ, Shieh M, Cosgrove DJ. 1995. Molecular cloning and sequence analysis of expansins—a highly conserved, multigene family of proteins that mediate cell wall extension in plants. *Proc Natl Acad Sci U S A.* 92:9245–9249.

Smant G, Stokkermans JP, Yan Y, et al. (13 co-authors). 1998. Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci U S A.* 95:4906–4911.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.

Vannerum K, Huysman MJ, De Rycke R, et al. (12 co-authors). 2011. Transcriptional analysis of cell growth and morphogenesis in the unicellular green alga *Micrasterias* (Streptophyta), with emphasis on the role of expansin. *BMC Plant Biol.* 11:128.

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. *Proc Natl Acad Sci U S A.* 110: 16444–16449.

Won H, Renner SS. 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proc Natl Acad Sci U S A.* 100:10824–10829.

Xi Z, Bradley RK, Wurdack KJ, Wong K, Sugumaran M, Bomblies K, Rest JS, Davis CC. 2013. Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* 13:227.

Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC. 2012. Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet.* 9:e1003265.

Yennawar NH, Li LC, Dudzinski DM, Tabuchi A, Cosgrove DJ. 2006. Crystal structure and activities of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize. *Proc Natl Acad Sci U S A.* 103:14664–14671.

Yoshida S, Maruyama S, Nozaki H, Shirasu K. 2010. Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328:1128.