RESEARCH ARTICLE   **Open Access**

Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles

André Busch[1], Etienne G. J. Danchin[2] and Yannick Pauchet[1*]

Abstract

Background: Cellulose, a major polysaccharide of the plant cell wall, consists of β-1,4-linked glucose moieties forming a molecular network recalcitrant to enzymatic breakdown. Although cellulose is potentially a rich source of energy, the ability to degrade it is rare in animals and was believed to be present only in cellulolytic microbes. Recently, it has become clear that some animals encode endogenous cellulases belonging to several glycoside hydrolase families (GHs), including GH45. GH45s are distributed patchily among the Metazoa and, in insects, are encoded only by the genomes of Phytophaga beetles. This study aims to understand both the enzymatic functions and the evolutionary history of GH45s in these beetles.

Results: To this end, we biochemically assessed the enzymatic activities of 37 GH45s derived from five species of Phytophaga beetles and discovered that beetle-derived GH45s degrade three different substrates: amorphous cellulose, xyloglucan and glucomannan. Our phylogenetic and gene structure analyses indicate that at least one gene encoding a putative cellulolytic GH45 was present in the last common ancestor of the Phytophaga, and that GH45 xyloglucanases evolved several times independently in these beetles. The most closely related clade to Phytophaga GH45s was composed of fungal sequences, suggesting this GH family was acquired by horizontal gene transfer from fungi. Besides the insects, other arthropod GH45s do not share a common origin and appear to have emerged at least three times independently.

Conclusion: The rise of functional innovation from gene duplication events has been a fundamental process in the evolution of GH45s in Phytophaga beetles. Both, enzymatic activity and ancestral origin suggest that GH45s were likely an essential prerequisite for the adaptation allowing Phytophaga beetles to feed on plants.

Keywords: Chrysomeloidea, Curculionoidea, GH45, Cellulase, Xyloglucanase, Horizontal gene transfer

Background

The major source of energy for most organisms on Earth is D-glucose. Through photosynthesis, plants have evolved the ability to biosynthesize organic D-glucose from inorganic carbon dioxide. Photosynthesis has provided plants with a nearly unlimited access to glucose, giving them a reservoir for storing energy as well as access to material for building structural components during plant growth. The plant cell wall (PCW) consists of several glucose-derived polysaccharides, which form a protective wall against biotic and abiotic stresses.

Traditionally, three kinds of polysaccharides are used as structural elements in the PCW: cellulose, hemicellulose and pectin. While the latter two are characterized by a variety of differently organized heteropolysaccharides, cellulose is a homopolymer and consists of β-1,4-linked anhydroglucose units forming a straight-chain polysaccharide. Through hydrogen bonding, individual chains attach to each other and form a resilient (para)crystalline structure [1]. On surface areas, cellulose is believed to organize itself into a state of low crystallinity, referred to as an "amorphous" state [2]. Depending on the developmental stage of the plant cell, the PCW is organized as follows: (i) the primary cell wall, which contains low amounts of crystalline cellulose (and surrounds plant cells in development) or (ii) the secondary cell wall, which

* Correspondence: ypauchet@ice.mpg.de
[1]Entomology, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany
Full list of author information is available at the end of the article

Busch *et al. BMC Evolutionary Biology* (2019) 19:100

Page 2 of 14

comprises large amounts of crystalline cellulose [3]. However, how native cellulose is organized in primary and secondary cell walls with regard to the ratio of amorphous to crystalline cellulose is still unclear [4, 5].

As the most abundant biopolymer on Earth [6], cellulose represents an abundant energy supply for any organism which has the ability to exploit it. Curiously, cellulose degradation has evolved only in few branches of the tree of life. Until the end of the twentieth century, cellulose degradation was only known to be performed by microorganisms such as plant pathogenic bacteria [7, 8], saprophytic fungi [9] or mutualistic symbionts in insects and ruminants [10, 11]. However, in 1998, the first endogenous cellulases of animal origin were identified in plant-parasitic nematodes [12] and termites [13]. Several other independent discoveries of cellulases in a variety of Metazoa followed, and to date endogenous cellulases encompass the phyla Arthropoda, Mollusca and Nematoda [14–18].

Cellulases are conventionally classified according to their mode of action. Endo-β-1,4,-glucanases (EC 3.2.1.4) break down cellulose by releasing randomly sized cellulose fragments and are known to act only on amorphous cellulose. Cellobiohydrolases (exo-β-1,4,-glucanases; EC 3.2.1.91) degrade cellulose from its terminal regions by releasing cellobiose and occasionally cellotriose. In microbes, cellobiohydrolases were shown to degrade amorphous as well as crystalline cellulose [19]. Finally, cellobiosidases (β-glucosidases; EC 3.2.1.21) accept the released cellobiose as substrate and convert it into glucose. All three types of cellulases act synergistically and are necessary to degrade the cellulosic network efficiently [20].

Cellulases are distributed into 14 of the 156 currently described families of glycoside hydrolases (GHs), according to the carbohydrate-active enzyme (CAZy) database (www.cazy.org) [21]. Assignment to different GH families is based on sequence similarities. The best-described cellulolytic GH families encompass GH5 [22] and GH9 [23]. Together with GH45s, GH5s and GH9s are found to be encoded by the genome of some insects [24–28]. Based on our previous work, GH45s are commonly distributed in the hyperdiverse Phytophaga clade of beetles [29], which encompasses the superfamilies Chrysomeloidea (leaf beetles and longhorned beetles) and Curculionoidea (weevils and bark beetles) [18, 30, 31]. With around 130,000 species distributed worldwide, which constitutes 50% of the plant feeding insect diversity, beetles of the Phytophaga clade represent the largest group of herbivorous insects on the planet. The first GH45 that was functionally characterized in a beetle originated from the mulberry longhorned beetle *Apriona germari* (Chrysomeloidea: Cerambycidae: Lamiinae), which had the ability to degrade amorphous cellulose [32]. Until recently, GH45s in beetles have been functionally characterized in only a few Chrysomeloidea species, mostly Cerambycidae

[31, 33, 34] and the western corn rootworm *Diabrotica virgifera virgifera* (Chrysomelidae: Galerucinae) [35], and another two from our previous study in the green dock leaf beetle *Gastrophysa viridula* (Chrysomelidae: Chrysomelinae) [36]. Although GH45 sequences have been identified in Curculionoidea beetles [18, 24, 28], to date none has ever been functionally characterized.

Interestingly, GH45s are not only found in multicellular organisms but are widely encoded by microbes [37–40]. The distribution of this gene family in the Metazoa appears to be patchy and has so far been recorded only in a few species within the phyla Mollusca [16, 41, 42], Nematoda [15, 43]; and Arthropoda [44]. If GH45s had evolved in the last common ancestor (LCA) of the Metazoa and subsequently been inherited by their descendants, we would expect the patchy distribution of GH45 genes observed within the Metazoa to be due to multiple independent losses. If this hypothesis were true, phylogenetic analyses would recover metazoan GH45s as a single monophyletic clade. However, two previous studies focusing on the evolutionary origin of GH45s in nematodes and mollusks have suggested instead that GH45s were acquired from a fungal donor by independent horizontal gene transfer events (HGT) [15, 16]. The first attempt to clarify the evolutionary history of GH45s in beetles also proposed an HGT from a fungal source but was unable to reach definite conclusions [45] because of the low number of sequences used for the phylogenetic analysis. A more comprehensive approach followed in 2014 [46], which included more GH45 sequences. Still, the variety of GH45 sequences in the latter study was poor, resulting in a similarly elusive outcome. Thus, the evolutionary history of GH45s appears to be complex, and their inheritance in beetles remains enigmatic.

Therefore, the major aim of our study was to trace the evolutionary origin of the GH45 family within the Phytophaga and to analyze how the function of the corresponding proteins evolved in this large group of beetles. Based on previous research on the ancestral origin of GH45s [15, 16], we hypothesize that an HGT event occurred at one or more stages of the evolution of the Phytophaga. Additionally, we analyzed other Arthropods, including Oribatida and Collembola, as well as several non-arthropod species, including Nematoda, Tardigrada and Rotifera. In this study, we combined functional and phylogenetic analyses to unravel the origin, evolution and functional diversification of the GH45 family in Phytophaga beetles. We first functionally characterized 37 GH45s from five beetle species -- four beetles of the Chrysomelidae (leaf beetles) and a beetle of the Curculionidae (weevils) -- to determine whether these GH45s harbor cellulase activity, and whether they may have evolved other functions. We then combined these functional data with amino acid alignments of the GH45

Busch *et al. BMC Evolutionary Biology* (2019) 19:100

Page 3 of 14

catalytic sites to pinpoint amino acid substitutions which might lead to substrate shifts. Finally, we performed phylogenetic analyses to assess (i) how many GH45 genes were present in the LCA of the Phytophaga and (ii) whether this gene family is ancestral in the Metazoa or, instead, acquired by HGT. The aim of our study was to provide the first comprehensive overview regarding the evolution in beetles of the GH45 family and to assess the role of these genes in the evolution of herbivory.

Results

Functional analyses of the Phytophaga GH45 proteins reveal distinct enzymatic characteristics

Our previous transcriptome analyses [18, 30, 46] revealed a set of endogenous GH45 genes distributed within several beetles of the superfamilies Chrysomeloidea and Curculionoidea. We investigated the product of GH45 genes from four beetle species belonging to the family Chrysomelidae, namely, *Chrysomela tremula* (CTR; Aspen leaf beetle), *Phaedon cochleariae* (PCO; mustard leaf beetle), *Leptinotarsa decemlineata* (LDE; Colorado potato beetle) and *Diabrotica virgifera virgifera* (DVI; Western corn rootworm), and from one species belonging to the family Curculionidae, the rice weevil *Sitophilus oryzae* (SOR; rice weevil), for a total of 33 GH45 sequences (Additional file 1: Table S1). By re-examining the corresponding transcriptomes as well as the recent draft genome of *L. decemlineata* [27], we identified four extra GH45 sequences (Additional file 1: Table S1). The resulting 37 GH45s were successfully expressed in *Sf*9 insect cells (Fig. 1a). All GH45s had an apparent molecular weight of ~ 35 kDa (Fig. 1a). The increase in molecular size compared to the expected size (~ 25 kDa) was likely due to post-translational N-glycosylations as well as to the artificially added V5/(His)$_6$-tag.

To explore the cellulolytic capabilities of these proteins, we first applied crude *Sf*9 culture medium containing individual recombinant GH45s to agarose plates supplemented with 0.1% carboxymethylcellulose (CMC) (Fig. 1b). Activity halos were visible for at least two GH45s per target species. The intensity of the observed activity halos varied from large clearing zones (for example, PCO4 or LDE2) to small or medium ones (such as PCO3 or LDE7). These differences were likely due to the catalytic efficiency of each individual GH45 as well as to the concentration of the crude protein extracts we used. Each clearing zone, independent of its intensity and size, indicated endo-β-1,4-glucanase activity.

To further assess enzymatic characteristics of these GH45s, we performed assays with a variety of plant cell wall-derived polysaccharides as substrates and analyzed the resulting breakdown products on thin layer chromatography (TLC) (Additional file 1: Figures S1 to S5). We were able to confirm the cellulolytic activity initially observed on CMC agar plates (CTR1 and CTR3, PCO1, PCO3, PCO4 and PCO6, LDE2, LDE3, LDE5, and LDE6, DVI1, DVI2 and DVI8–11, SOR1 and SOR2). Each of these enzymes was able to break down CMC, regenerated amorphous cellulose (RAC) and cellulose oligomers. Interestingly, LDE10 did not show activity against cellulosic polymers but preferentially degraded cellopentaose and cellohexaose (Additional file 1: Figure S3). Similarly, but with much weaker efficiency, PCO5 degraded cellohexaose (Additional file 1: Figure S2). Together with the plate assays, our TLC analyses clearly demonstrated that beetle-derived GH45s processed cellulosic substrates using an endo-active mechanism, which suggests that these enzymes are endo-β-1,4-glucanases. Several cellulolytically active GH45s derived from the four leaf beetle species displayed additional activity towards the hemicellulose glucomannan, for example, CTR1, PCO3 and LDE2 (Additional file 1: Figures S1 to S3). LDE7 exhibited the highest enzymatic activity against glucomannan, whereas its activity against amorphous cellulose substrates could be visualized only by plate assay.

Interestingly, TLC allowed us to detect several enzymes (CTR2, PCO7, LDE5, LDE11, DVI5, DVI6 and SOR3 to SOR5) which were able to degrade xyloglucan instead of cellulose (Additional file 1: Figures S1 to S5). The size of the resulting breakdown products seemed to correlate with heptamers and octamers, indicating that these proteins were endo-β-1,4-xyloglucanases; however, the actual size of the resulting breakdown products is difficult to assess because some glucose moieties that make up the backbone of xyloglucan are substituted with xylose residues. LDE5 displayed activity towards amorphous cellulose substrates (Fig. 1b and Additional file 1: Figure S3), and is thus, according to our data, the only example of a beetle-derived GH45 able to degrade xyloglucan as well as amorphous cellulose. Additionally, each of the 37 GH45s was tested against xylan and no activity was detected (data not shown).

The Chrysomelid-derived PCO8, LDE1, LDE4, LDE8, LDE9, DVI3, DVI4 and DVI7 did not exhibit activity towards any of the substrates used in this study. We wondered whether substitutions of catalytically important residues may have caused their apparent loss of activity. To test this hypothesis, we performed an amino acid alignment of target beetle GH45 sequences, including a fungal GH45 sequence, as a reference for which the structure has been resolved [37]; then we screened for amino acid substitutions and compared these to the reference fungal sequence (Fig. 2). According to Davies et al. [37], both the proton donor (catalytic acid) and the acceptor (catalytic base) of the catalytic dyad should be aspartates (Asp10 and Asp121). In LDE9, the catalytic base was substituted for an asparagine, whereas in DVI4, the catalytic acid was substituted for a valine (Fig. 2). In
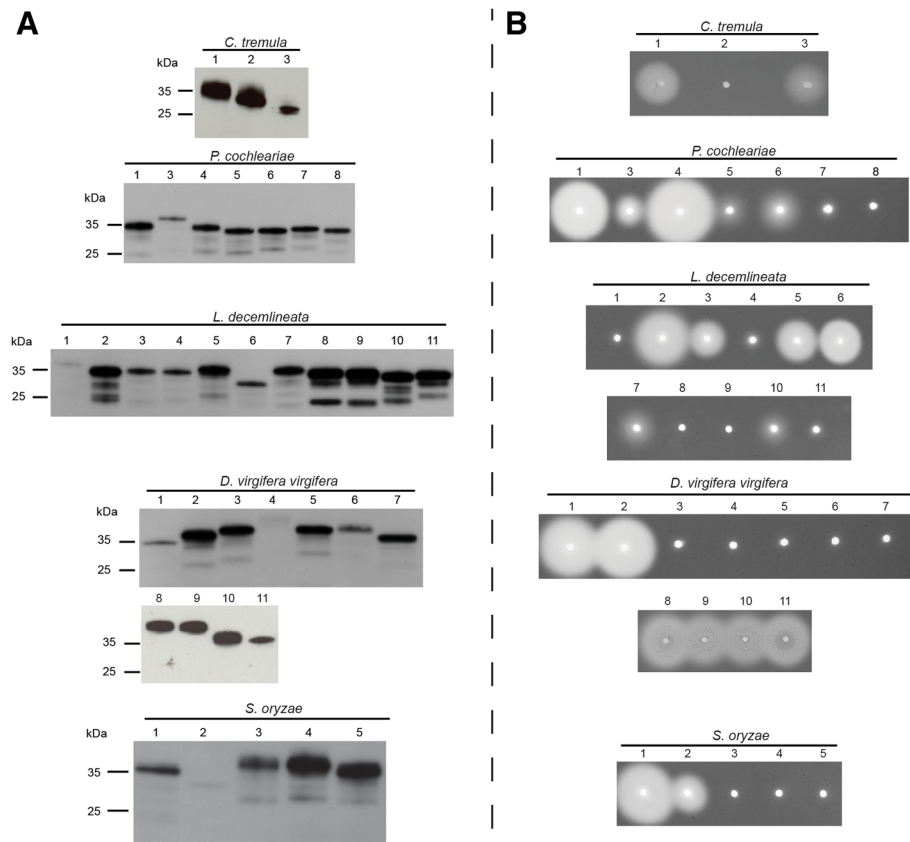
Fig. 1 Western blot and CMC-based agarose-diffusion assay of target GH45 proteins. **a**) Western blot of target recombinant enzymes expressed in frame with a V5/(His)₆ after heterologous expression in insect *Sf*9 cells. After 72 h, crude culture medium of transfected cells was harvested and analyzed by Western blotting using an anti-V5 HRP-coupled antibody. **b**) Crude culture medium of transfected cells was applied to an agarose-diffusion assay containing 0.1% CMC. Activity halos were revealed after 16 h incubation at 40 °C using Congo red. Numbers above Western blot and agarose-diffusion assays correspond to the respective species of GH45s depicted in Additional file 1: Table S1

both cases, the loss of the carboxyl unit of the functional group likely caused the proteins to lose the catalytic activity. No substitution event of the catalytic residues was observed for PCO8, LDE1, LDE4, LDE8, DVI3, and DVI7. Thus, we decided to investigate several conserved sites known to affect the enzymatic activity of GH45s [37]. In addition, we investigated three other sites crucial for enzymatic activity: (i) a proposed stabilizing aspartate (Asp114), (ii) a conserved alanine (Ala74) and (iii) a highly conserved tyrosine (Tyr8) (Fig. 2). Apart from two substitution events of Tyr8 in LDE9 and LDE3, this amino acid remained conserved in all other beetle GH45 sequences. LDE9 already possessed a mutation in its catalytic acid, which was likely responsible for the lack of activity. In LDE3, a substitution from Tyr8 to Phe8 did not significantly impact the catalytic abilities of this protein, likely because the side-chains of both amino acids are highly similar and differ only in a single hydroxyl group. When examining the proposed stabilizing

site Asp114, we observed several amino acid substitutions that correlated with a loss of activity in PCO8, DVI3, DVI4, LDE1 and LDE8. Amino acid changes at the Asp114 position were also observed in PCO3 and CTR1, but were not correlated with a loss of enzymatic activity. Since LDE4 and DVI7 appeared to have no mutation in Asp114, we screened the Ala74 residue for substitutions; the amino acid exchange we observed, from alanine to glycine in both cases, may have caused the loss of activity in these two proteins. Altogether, amino acid substitutions at important sites could be detected in some apparently inactive GH45s, but not in all of them. It may be that the proteins for which we did not find amino acid substitutions are still active enzymes, and we have just not yet found the right substrate; alternatively, we have not yet checked all the amino acid positions, some of which could also be crucial for catalysis.

Interestingly, all Chrysomelid GH45 xyloglucanases (except LDE5 and DVI6), including *G. viridula* GVI1

GH45 enzymatic activity	Proton Acceptor (Tyr8, Asp10)	Proton Donor (Asp121)	Stabilizer (Asp114)	conserved Alanine (Ala74)
HIN1 — green	STRYWDC	HFDLN	GDL	FAAT
PCO1 — green	TTRYWDC	QFDLA	GDL	FAAG
PCO3 — green, yellow	TTRYWDC	QFDIE	GPL	FAAA
PCO4 — green, yellow	TTRYWDC	QFDLA	SDL	FSAA
PCO5 — green, yellow	TTRYWDC	HFDLA	SDL	FSAA
PCO6 — green, yellow	TTRYWDC	HFDLA	SDL	FSAA
PCO7 — blue	TTRYWDC	LFDIA	PET	FVAA
PCO8 — red	TTRYWDC	QFDLQ	SPL	FAAA
CTR1 — green, yellow	TTRYWDC	QFDIE	SPL	FAAA
CTR2 — green, yellow	TTRYWDC	LFDIA	PES	FAAA
CTR3 — green, yellow	TTRYWDC	QFDLA	SDL	FSAA
LDE1 — red	TTRYWDC	QFDIE	SPL	FSAV
LDE2 — green, yellow	TTRYWDC	HFDIA	SDL	FSAS
LDE3 — green, yellow	TTRFWDC	QFDIA	SDL	YVAA
LDE4 — red	TTRYWDC	QFDLA	ADA	FAGA
LDE5 — green, blue	TTGYWDC	QFDLA	ADL	FTAA
LDE6 — green	TTRYWDC	QFDIA	GDL	FAAA
LDE7 — green, yellow	TTRYWDC	QFDIA	SDL	YVAA
LDE8 — red	TTRYWDC	QFDLA	SPL	FSAA
LDE9 — red	TTRHWNC	QFDLA	TAV	FSGA
LDE10 — green	STLYWDC	HFDLA	SDL	FSAS
LDE11 — blue	TTRYWDC	MFDIA	PEE	FVAA
DVI1 — green	TTRYWDC	HFDIA	SDL	FGAA
DVI2 — green	TTRYWDC	QFDIA	GDL	FAAV
DVI3 — red	TTRYWDC	QFDLG	EEY	FASA
DVI4 — red	TSRYWDC	EFVIA	TNF	FAAA
DVI5 — blue	TTRYWDC	LFDIA	PET	YVAA
DVI6 — blue	TTRYWDC	EFDIA	PHA	FVAA
DVI7 — red	TTRYWDC	HFDLQ	ADA	FAGV
DVI8 — green, yellow	TTRYWDC	QFDLQ	SDL	WVAA
DVI9 — green, yellow	TTRYWDC	HFDIQ	EAL	FVAA
DVI10 — green, yellow	TTRYWDC	QFDLQ	SDL	FAAV
DVI11 — green, yellow	TTRYWDC	HFDIQ	EAL	FVAA
SOR1 — green	TTRYWDC	HFDIQ	GDL	FAAA
SOR2 — green	TTRYWDC	HFDIQ	ADL	FAAA
SOR3 — blue	TTRYWDC	NIEIA	SSV	FAAA
SOR4 — blue	TSRYWDC	NIEIG	SSV	FAAA
SOR5 — blue	TSRYWDC	NIEVA	SNS	FAAA

Fig. 2 GH45 amino acid alignment of the catalytic residues. We used a GH45 sequence of *Humicola insulens* (HIN1) as a reference sequence (Accession: 2ENG_A) [37]. According to HIN1, we chose to investigate the catalytic residues (ASP10 and ASP121) as well as a conserved tyrosine (TYR8) of the catalytic binding site, a crucial substrate-stabilizing amino acid (ASP114) and an essential conserved alanine (ALA74). Arrows indicate amino acid residues under investigation. If the respective amino acid residue is highlighted in green, it is retained in comparison to the reference sequence; otherwise it is highlighted in red. GH45 enzymatic activity was color-coded based on the respective substrate specificity (green dots = endo-β-1,4-glucanase, blue dots = endo-β-1,4-xyloglucanase, yellow dots = (gluco)mannanase, red dots = no detected activity)

from our previous study, [36], displayed a substitution from aspartate to glutamate at the stabilizing site (Asp114) (Fig. 2). As glutamate differs from aspartate only by an additional methyl group within its side chain, we believe that this exchange may have contributed to the substrate shift. Interestingly, and in contrast to Chrysomelidae-derived GH45 xyloglucanases, we found that GH45 xyloglucanases from the Curculionidae *S. oryzae* (SOR3 to SOR5) had a substitution from aspartate to glutamate in the proton donor residue (Asp121). We also believe that, in *S. oryzae*, this particular substitution may have contributed to the preference for xyloglucan over cellulose as a substrate.

In summary, we demonstrated that each species investigated encoded at least two cellulolytic GH45s that are able to degrade amorphous cellulose. We also demonstrated that at least one GH45 per species evolved the ability to specifically degrade xyloglucan. Interestingly, several GH45s did not show activity on any of the substrates we tested, suggesting that they have become pseudo-enzymes or are active on substrates not tested here.

Phylogenetic analyses reveal multiple origins of GH45 genes during the evolution of Metazoa

We used phylogenetic analyses to reconstruct and gain further insight into the evolutionary history of beetle-derived GH45 genes. To achieve this goal, we collected amino acid sequences of GH45s available as of February 2018, including those from the CAZy database (http://www.cazy.org) [21] as well as from several transcriptome datasets accessible at NCBI Genbank. Interestingly, we realized that the presence of GH45 genes in arthropods was not restricted to Phytophaga beetles: these genes were also distributed in transcriptomes/genomes of species of springtails (Hexapoda: Collembola) and of species of Oribatida mites (Arthropoda: Chelicerata) (Additional file 1: Table S2). In addition, we identified GH45 sequences in two other groups of Metazoa, namely, tardigrades and rotifers. Notably, our homology search did not retrieve any mollusk-derived GH45s; given the distant similarity of these GH45s to any of those we investigated, this absence is not surprising. The patchy distribution of GH45 genes throughout the arthropods and, more widely, the Metazoa, could be due to either the presence of GH45 genes in a common ancestor, followed by multiple gene losses, or from multiple independent acquisitions from foreign sources (i.e., HGT). To test these hypotheses, we collected a diverse set of GH45 sequences of microbial and metazoan origins resulting in 264 sequences (Additional file 1: Table S2). Subsequently, redundancy at 90% identity level between sequences was eliminated, resulting in 201 non-redundant GH45 sequences.

According to both Bayesian and maximum likelihood phylogenies (Additional file 1: Figures S3, S6 and S7), the arthropod-derived GH45 sequences were not monophyletic but globally fell into three separate groups. One highly supported monophyletic clade (posterior probability (PP) =1.0, bootstrap =85) grouped all the Phytophaga beetle GH45 sequences. This clade was most closely related to a group of Saccharomycetales fungi (PP = 0.88, bootstrap = 44). Then this clade branched to yet another group of Saccharomycetales Fungi (PP = 1.0, bootstrap = 56). A second monophyletic clade grouped all the Oribatida mites GH45 sequences, although with moderate support on the branch (PP = 0.72, bootstrap = 3). Finally, a third clade (PP = 0.96, bootstrap = 14) grouped all the GH45 sequences from Collembola. This last group was not monophyletic: a bacterial GH45 sequence was interspersed within this clade and separated the Collembola GH45 sequences into two subgroups. In addition to the arthropods, the nematode GH45 sequences formed a highly supported monophyletic clade (PP = 1.0, bootstrap = 84) which was connected to a clade of fungal-derived sequences. This connection was highly supported (PP = 0.93, bootstrap = 69). The two other groups of Metazoa (tardigrades and

rotifers) were located in a separate clade with species of Neocallimastigaceae fungi (Chytrids), (PP = 0.94, bootstrap = 14). Overall, this analysis showed that neither arthropods nor, more generally, metazoan GH45 sequences, originated from a common ancestor, as they were scattered in multiple separate clades rather than forming a monophyletic metazoan clade.

In summary, our phylogenetic analyses illustrated that the evolutionary history of GH45s in the Metazoa was complex and pointed to the possibility that this gene family evolved several times independently in multicellular organisms. More specifically, our analyses suggested that this gene family had evolved at least three times independently in arthropods. Finally, our data pointed toward an acquisition of GH45 genes by the LCA of Phytophaga beetles -- presumably through an HGT event -- from a fungal donor.

The structure of GH45 genes in Phytophaga beetles supports a single origin before the split of the Chrysomeloidea and Curculionoidea

The monophyly of the Phytophaga-derived GH45s in the above phylogenetic analyses suggests a common ancestral origin in this clade of beetles. If the presence of a GH45 in the Phytophaga beetles had resulted from a single acquisition in their LCA, we hypothesized that the GH45 genes present in current species of leaf beetles, longhorned beetles and weevils would share a common exon/intron structure. To test this hypothesis, we mined the publicly available genomes of three species of Curculionidae, including *Hypothenemus hampei* (coffee berry borer) [28], *Dendroctonus ponderosae* (mountain pine beetle) [24] and *S. oryzae* (unpublished), as well as the genomes of the Chrysomelidae *L. decemlineata* [27] and of the Cerambycidae *Anoplophora glabripennis* (Asian longhorned beetle) [25]. We were able to retrieve the genomic sequence corresponding to each of the GH45 genes present in these beetle species, with the exception of DPO9, which we did not find at all, and SOR3 and SOR4, which we were able to retrieve only as partial genomic sequences. Our results showed that the number of introns varied between the different species (Additional file 1: Figure S8). In *L. decemlineata* (representing Chrysomelidae), we identified a single intron in each of the GH45 genes (except for LDE11, which had two). For *A. glabripennis* (representing Cerambycidae), we found two introns in each of the two GH45 genes. In *H. hampei*, *D. ponderosae* and *S. oryzae* (all representing Curculionidae), the number of introns ranged from three to five. Interestingly, all GH45 genes in these five species possessed an intron placed within the part of the sequence encoding the predicted signal peptide. Apart from DPO7 and DPO8, these introns were all in phase one. This gene structure of Chrysomelid- and Curculionid-derived GH45

genes correlated well with our previous study investigating the gene structure of PCW-degrading enzymes, including GH45 genes, in the leaf beetle *Chrysomela tremula* [26]. The conservation of the phase and the position of this intron indicated that the LCA of the Phytophaga likely possessed a single GH45 gene having a phase one intron located in a part of the sequence encoding a putative signal peptide. To assess whether that particular intron is also present in the most closely related fungal species, we blasted the genomes of Saccharomycetaceae and Neocallimastigaceae fungi (NCBI, whole-shotgun genome database) using the protein sequence of GH45–1 of *C. tremula* as a query. We did not detect any introns in fungal GH45 sequences (data not shown), suggesting that the intron was acquired after the putative HGT event, in an ancestor of Phytophaga beetles. The diversity of the overall intron-exon structure in phytophagous beetles likely resulted from subsequent and independent intron acquisition. In summary, and together with the monophyly of beetle-derived GH45s (Fig. 3), our analysis highly supports a common ancestral origin of beetle GH45.

Evolution of the GH45 family after the initial split of Chrysomeloidea and Curculionoidea

We mined publicly available transcriptome and genome datasets of Phytophaga beetles (Additional file 1: Table S3) and collected as many GH45 sequences as possible. We curated a total of 266 GH45 sequences belonging to 42 species of Phytophaga beetles. After amino acid alignment, we decided to exclude 60 partial GH45 sequences from our phylogenetic analysis because these were too short. We performed a "whole Phytophaga" phylogenetic analysis on the remaining 206 curated GH45 sequences using maximum likelihood (Fig. 4). Deep nodes were very poorly supported indicating a high gene birth/death dynamics for this family of proteins in during the course of evolution of Phytophaga beetles. On the other hand, some nodes were highly supported and we will detail them below. It is also interesting to note that, our phylogenetic analysis indicated that no evident orthologous GH45 genes could be defined between species of Chrysomeloidea and Curculionoidea (Fig. 4).

Clade 'a' comprised Brentidae- and Curculionidae-derived GH45s including SOR3 to SOR5 (Fig. 4). According to our functional data, SOR3 to SOR5 act as xyloglucanases, suggesting that other GH45 proteins present in clade 'a' may have also evolved to degrade xyloglucan. To support this hypothesis, we compared the catalytic residues of SOR3–5 to those of the other Curculionidae and Brentidae-derived GH45 sequences from clade n (Additional file 1: Figure S9). We detected substitutions from an aspartate to a glutamate at Asp121 in all Curculionidae-derived GH45 sequences of this clade but not in the Brentidae-derived sequences, suggesting that Curculionidae-derived GH45s of this clade

were likely to possess xyloglucanase activity. Functional analyses of the Brentidae-derived sequences present in clade 'a' will be needed to determine whether these proteins are also xyloglucanases or whether they fulfill another function.

GH45s of clade 'r', with a bootstrap support of 84, contained only Curculionidae-derived sequences (Fig. 4). Within this clade, we found SOR1 and SOR2, which are, according to our functional data, endo-active cellulases. Their presence in clade 'r' implies that other GH45s of this cluster exhibit potential endo-cellulolytic activity. To further support this hypothesis, we again investigated amino acid residues of the catalytic site by comparing SOR1 and SOR2 to other GH45 sequences in this clade (Additional file 1: Figure S9). We did not find crucial substitutions in any of the investigated sites, implying that all GH45 proteins of this clade may have retained endo-β-1,4-glucanase activity. GH45 sequences present in the other Curculionidae and Brentidae-specific clades did not harbor any amino acid substitutions which could impair their catalytic properties, and they may all possess the ability to break down amorphous cellulose. More functional analyses will be necessary to assess the function of these proteins.

Regarding Chrysomeloidea-derived sequences, a highly supported clade (Fig. 4, clade 'b') contained GH45 sequences of two subfamilies of Chrysomelidae, namely, Chrysomelinae and Galerucinae. Our functional analyses revealed that this clade contained GH45 proteins possessing xyloglucanase activity, including DVI5 and DVI6 from *D. vir. Virgifera*. The remaining Galerucinae-derived GH45s, such as those from the Alticines *Phyllotreta armoraciae* and *Psylliodes chrysocephala*, present in clade 'b' have yet to be functionally characterized. When the catalytic residues of active xyloglucanases from our study were compared to the uncharacterized GH45 sequences present in clade 'b', we observed that at least PAR6 and PCH5 of the Galerucinae and OCA10 and CPO2 of the Chrysomelinae had congruent substitutions (ASP114 > Glu114), which likely enabled those proteins to also degrade xyloglucan (Additional file 1: Figure S10). Therefore, it is highly likely that the LCA of the Chrysomelinae and the Galerucinae possessed at least two GH45 proteins, an endo-acting cellulase and a xyloglucanase.

Clade 'm' consisted solely of Lamiinae-derived sequences and in fact encompassed all Cerambycidae-derived GH45s identified to date (Fig. 4). Several of those GH45s had been previously functionally characterized as cellulases including AJA1 and AJA2 [31], AGE1 and AGE2 [32, 47], AGL1 and AGL2 [25], ACH1 [34] and BHO1 [33]. Investigating their catalytic residues revealed no critical amino acid substitutions (Additional file 1: Figure S10), indicating that each yet-uncharacterized Cerambycidae GH45s from clade 'm'
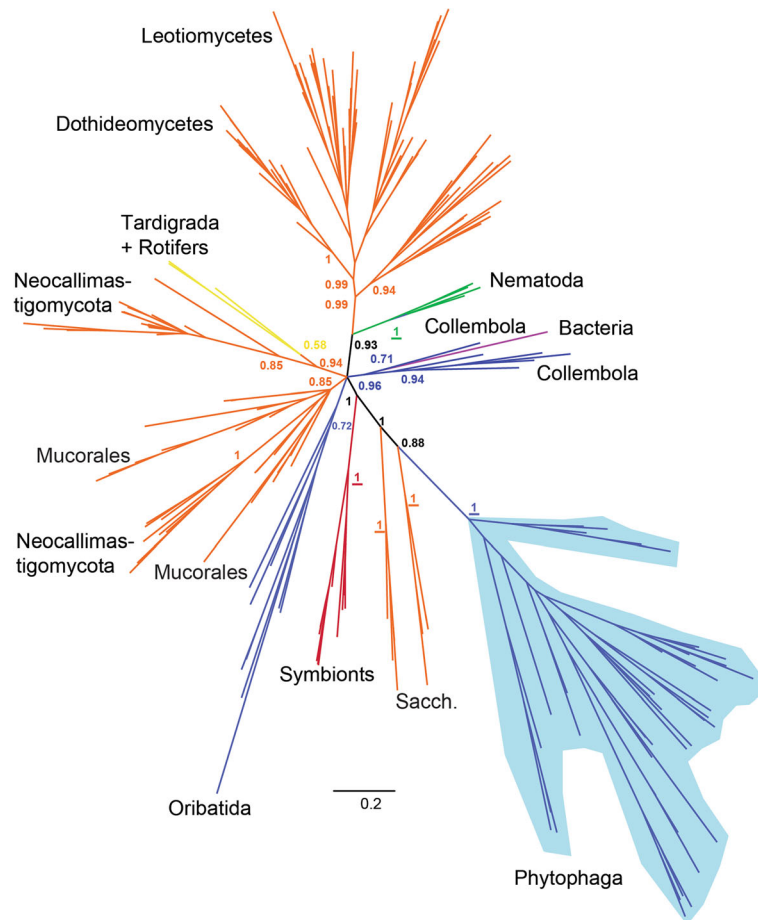
Fig. 3 Global phylogeny encompassing GH45 proteins from various taxa. Bayesian-based phylogenetic analysis of GH45 sequences. 264 GH45 sequences of microbial and metazoan origin were initially collected (see Methods), and their redundancy was eliminated at 90% sequence similarity, resulting in a total of 201 sequences. Posterior probability values are given at crucial branches. If values are depicted in bold, the same branch appeared in the corresponding maximum likelihood analysis (see Additional file 1: Figures S6 and S7). If underlined, the maximum likelihood node was highly supported (bootstrap values > 75). Detailed sequence descriptions including accession numbers are given in Additional file 1: Table S2. Arthropoda are represented in blue, fungi in orange, protists in red, bacteria in purple, Nematoda in green and other Metazoa in yellow. Sacch. = Saccharomycetales fungi; Neocallim. = Neocallimastigaceae fungi

(i.e. OCH1, MMY1, PHI and MAL1) may also possess cellulolytic activity.

In summary, our focus on Phytophaga-derived GH45s with regards to enzymatic characterization and ancestral origin allowed us to postulate that at least one GH45 protein was present in the LCA of the Phytophaga beetles and that this GH45 protein likely possessed cellulolytic activity. After the split between Chrysomeloidea and Curculionoidea, the GH45 gene family evolved through gene duplications at the family, subfamily and even genus/species level. Finally, according to our data, the ability of these beetles to break down xyloglucan, one of the major components of the primary plant cell wall, evolved at least twice, once in the LCA of the Chrysomelinae and Galerucinae and once in the LCA of the Curculionidae.

Discussion

In our previous research, we found that several beetles of the Phytophaga encoded a diverse set of GH45 putative cellulases [18]. Here, we demonstrated that in each of the five Phytophaga beetles investigated, at least two of these GH45s possess cellulolytic activity. This discovery is in accordance with other previously described GH45 proteins from Insecta [31], Nematoda [15], Mollusca [42], Rotifera [48] and microbes [49].

Surprisingly, several GH45 proteins were able to degrade glucomannan in addition to cellulose. We hypothesize that GH45 bi-functionalization may have occurred as a result of the chemical similarities between cellulose and glucomannan. Glucomannan is a straight chain polymer consisting of unevenly distributed glucose and mannose moieties. GH45 cellulases could recognize two adjoining glucose moieties

Fig. 4 Phylogenetic relationships of Phytophaga-derived GH45s. A maximum-likelihood-inferred phylogeny of the predicted amino acid sequences of beetle-derived GH45s was performed. Bootstrap values are indicated at corresponding branches. Information on sequences and their accession number are given in Additional file 1: Table S2. Dots indicate GH45s characterized to date and are color-coded based on their activity: green = endo-β-1,4-glucanases; blue = endo-β-1,4-xyloglucanases; yellow = (gluco)mannanases; red = no activity detected. Color coding in reference to the respective subfamily of Curculionoidea: pink = Scolytinae (Curculionidae); brown = Entiminae (Curculionidae); purple = Cyclominae (Curculionidae); gray = Curculioninae (Curculionidae); yellow = Molytinae (Curculionidae); light blue = Brentinae (Brentidae); dark blue = Dryophthorinae (Curculionidae). Color coding in reference to the respective subfamily of Chrysomeloidea: dark green = Chrysomelinae (Chrysomelidae); light green = Galerucinae (Chrysomelidae); orange = Lamiinae (Cerambycidae); cyan = Cassidinae (Chrysomelidae)

in the glucomannan chain, thus allowing hydrolysis to occur. Notably, enzymes specifically targeting mannans of the PCW are rare in Phytophaga beetles. So far, they have been identified and characterized only in *G. viridula* and *Callosobruchus maculatus* (GH5 subfamily 10 or GH5_10) [18, 50], and one GH5_8 has been characterized in the coffee berry borer *H. hampei* [51]. But, in contrast to the activity on glucomannan of some GH45s we observed here, those GH5_10s and GH5_8 were true mannanases, displaying activity towards galactomannan as well as glucomannan. Although our experiments suggested some GH45 cellulases were also active on glucomannan, we believe that the activity these proteins carry out could be important for the degradation of the PCW in the beetle gut. In fact, mannans, including glucomannan, can make up to 5% of the plant primary cell wall [52] and may be a crucial enzymatic target during PCW degradation. This hypothesis is further supported by the presence of at least one GH45 protein

with some ability to degrade glucomannan in each of the Chrysomelid beetles for which we have functional data.

Another interesting discovery was that several GH45 proteins have lost their ability to use amorphous cellulose as a substrate and evolved instead to degrade xyloglucan, the major hemicellulose of the plant primary cell wall [53]. We believe that the initial substrate shift from cellulose to xyloglucan has likely been promoted by similarities between the substrate backbones (in both cases β-1,4 linked glucose units). The major difference between cellulose and xyloglucan is that the backbone of the latter is decorated with xylose units (which in turn can be substituted by galactose and/or fucose). We presume that the substrate shift from a straight chain polysaccharide such as cellulose to a more complex one such as xyloglucan requires the similar complex adaptation of the enzyme to its novel substrate. However, in contrast to glucomannan-degrading GH45s, GH45 xyloglucanases have apparently completely lost their ability

to use amorphous cellulose as a substrate. Here, we clearly demonstrated that, following several rounds of duplications, GH45s in Chrysomelid beetles have evolved novel functions in addition to their ability to break down amorphous cellulose, allowing these insects to degrade two additional major components of the PCW, namely. Glucomannan and xyloglucan. This constitutes yet another clear example of functional diversification mediated by gene duplication [54] giving rise to a whole multigene family in Phytophaga. This broadening of their functions further emphasizes that GH45 proteins may have likely been an important innovation during the evolution of the Phytophaga beetles and may have strongly contributed to their radiation. In summary, the ability of GH45 proteins to degrade a variety of substrates either as monospecific or as bi-functionalized enzymes indicates that these proteins are particularly prone to substrate shifts.

According to our data, the ability to break down xyloglucan using a GH45 protein has evolved at least twice independently in Phytophaga beetles, once in the LCA of Chrysomelinae and Galerucinae and once in the LCA of the Curculionidae or of the Curculionidae and Brentidae. Once the first Brentidae-derived GH45s are functionally characterized, we will know more. Given that genome/transcriptome data for a majority of families and subfamilies are lacking throughout the Phytophaga clade, we expect that other examples of independent evolution of GH45 xyloglucanases will be revealed in the future. It is important to note that the ability to degrade xyloglucan, which represents an important evolutionary innovation for Phytophaga beetles, may not be linked solely to the evolution of the GH45 family. In fact, in A. glabripennis (Cerambycidae: Lamiinae), a glycoside hydrolase family 5 subfamily 2 (GH5_2) protein has evolved to degrade xyloglucan; additionally, orthologous sequences of this GH5_2 xyloglucanase have been found in other species of Lamiinae [25].

The ability of GH45s to break down xyloglucan correlated with a substitution event from an aspartate to a glutamate residue at a stabilizing site (Asp114) within the Chrysomelidae. Interestingly, the same amino acid exchange was present in SOR3-SO5 but was located at the catalytic acid (Asp121) rather than the stabilizing site (Asp114). Aspartate and glutamate share the same functional group but differ in the length of their side chain. Thus, the preservation of the functional group coupled with an elongated side chain has likely contributed to the substrate switch of those GH45 proteins which when turned on allows xyloglucan to be degraded. Notably, DVI6 and LDE5 do not share that particular substitution but are able to degrade xyloglucan. Therefore, we believe that the transition from cellulase to xyloglucanase has not been driven solely by a single amino acid substitution, but has been triggered by changes at other positions.

According to the carbohydrate-active enzyme (CAZy) database [21], GH45s encompass 385 sequences (as of February 2018) distributed throughout fungi, bacteria and Metazoans. Interestingly, the distribution of GH45s within Metazoans is rather patchy, encompassing to date only one clade of Nematodes [15, 43], some Arthropods [17, 18], bdelloid Rotifers [48] and some Mollusks [16]. To date, in insects, GH45s are restricted to Phytophaga beetles. We searched several other arthropod genome/transcriptome datasets, including beetles other than Phytophaga, and all publicly available insect genomes, as well as publicly available genomes of Collembola and Oribatida mites. Except for the latter two, we were unable to retrieve GH45 sequences from other arthropods. The patchy distribution of GH45 sequences among Metazoa suggests that these proteins were either acquired multiple times throughout animal evolution or that massive differential gene loss occurred within multicellular organisms. Surprisingly, our phylogenetic analyses clearly showed that the arthropod-derived GH45s, rather than clustering together, formed three separate monophyletic groups. In fact, all metazoan-derived GH45s clustered separately, forming independent monophyletic groups. Hence, the "multiple loss" hypothesis appears to be less parsimonious as it implies the existence of multiple GH45s in the LCA of Ophistokonta (Fungi and Metazoa) followed by reciprocal differential gene losses and multiple independent total gene losses in many animal lineages. Intriguingly, the closest clade to the Phytophaga GH45 sequences was composed of fungal GH45s. Our phylogenetic analyses could not identify a specific donor species/group but both suggested species of Saccharomycetales or Neocallimastigaceae fungi as potential source. The most parsimonious explanation for the appearance and current distribution of GH45 genes in Phytophaga beetles is that at least one gene was horizontally acquired from a fungal donor. A similar scenario may have been responsible for the presence of GH45 genes in Oribatida and Collembola, but this hypothesis remains speculative until more sequences from both these orders are identified. In addition to the monophyly of Phytophaga-derived GH45 sequences, a common origin was further suggested by the fact that the position and the phase of the first intron was (except for two cases) conserved across GH45 genes from the species of Cerambycidae, Chrysomelidae and Curculionidae for which genome data are available. If our hypotheses are correct, the LCA of all Phytophaga beetles most likely acquired a single GH45 gene from a fungal donor. As we did not find any fungal (Saccharomycetales or Neocalimastigaceae) introns corresponding to the proposed original intron, it appears that beetle-derived GH45 genes have acquired an intron after the HGT. The GH45 gene then likely underwent several duplications before the separation of the different Phytophaga clades, and

Busch *et al. BMC Evolutionary Biology*   (2019) 19:100

Page 11 of 14

these duplications continued independently after the diversification of this hyper-diverse clade of beetles.

Strikingly, nematode-derived GH45 sequences were consistently grouped together with a group of Saccharomycetales fungi distinct from the Phytophaga-related one, in each analysis we ran, clearly demonstrating that the closest relatives to their GH45 genes were fungal and from different fungi than the insect relatives. The origin of nematode-derived GH45 genes has been investigated, and their acquisition by HGT from a fungal source has been proposed [15, 43]. Here we provide the third independent confirmation of this fact.

Conclusion

Our research indicated that the Phytophaga GH45s have adapted to substrate shifts. In addition to cellulose, this adaptation led to the recognition and catalysis of two additional substrates, neither of which can be enzymatically addressed by any other GH family that those insects encode. Beetles of the Chrysomelidae have evolved to break down three components of the PCW (cellulose, xyloglucan and glucomannan) by using only GH45s. In concert with GH28 pectinases encoded by each investigated species [55], these beetles have evolved a near-complete set of enzymatic tools with which to deconstruct the PCW, allowing them to gain access to the nutrient-rich plant cell contents; in addition, PCW-derived polysaccharides are a potential source of energy. Our data also suggest that GH45 is not a Metazoan ancestral gene family but was likely acquired by through an HGT event from a fungal source in the LCA of the Phytophaga. The causes of the extraordinarily successful radiation of Phytophaga beetles are widely debated, but the evolution of specialized trophic interactions with plants is assumed to have played an important role. Both, enzymatic activity and ancestral origin suggest that GH45s, and other families of plant cell wall degrading enzymes, were likely an essential prerequisite for the adaptation allowing Phytophaga beetles to feed on plants.

Methods

Production of recombinant GH45 proteins

Open reading frames (ORFs) were amplified from cDNAs using gene-specific primers based on previously described GH45 sequences of *C. tremula*, *P. cochleariae*, *L. decemlineata*, *D. virgifera virgifera* and *S. oryzae* [18]. If necessary, full-length transcript sequences were obtained by rapid amplification of cDNA ends PCR (RACE-PCR) using RACE-ready cDNAs as described by [18]. For downstream heterologous expression, ORFs were amplified using a forward primer designed to include a Kozak sequence and a reverse primer designed to omit the stop codon. cDNAs initially generated for the RACE-PCRs, as described by [18], were used as PCR template, and the PCR reactions were performed using a high-fidelity Taq polymerase (AccuPrime, Invitrogen). Cloning into the pIB/V5-His TOPO (Invitrogen) vector, transfection in *Sf*9 insect cells and harvesting of crude recombinant protein extracts were essentially performed according to Busch et al. [36].

Enzymatic characterization

The enzymatic activity of recombinant proteins was initially tested on agarose diffusion assays using carboxymethylcellulose (CMC) as a substrate. Agarose (1%) plates were prepared, containing 0.1% CMC in 20 mM citrate/phosphate buffer pH 5.0. Small holes were made in the agarose matrix using cut-off pipette tips, to which 10 μl of the crude culture medium of each expressed enzyme was applied. After incubation for 16 h at 40 °C, activity was revealed by incubating the agarose plate in 0.1% Congo red for 1 h at room temperature followed by washing with 1 M NaCl until pale halos on a red background were visible. To investigate GH45 enzymatic activity in more detail, we analyzed their enzymatic breakdown products using thin layer chromatography (TLC). For that, the culture medium of transiently transfected cells was dialyzed and desalted as described in Busch et al. [50]. The following substrates were tested: CMC, Avicel, glucomannan, galactomannan and xyloglucan (all from Megazyme) with a final concentration of 0.5%. We also tested regenerated amorphous cellulose (RAC), prepared according to Zhang et al. [56]. Additionally, we used the cello-oligomers D-(+)-biose to D-(+)-hexaose (all from Megazyme), as substrates at a final concentration of 250 ng/μl. Samples were incubated and analyzed as previously described [50]. The reference standard contained 2 μg of each oligomer: glucose, cellobiose, cellotriose, cellotetraose and cellopentaose as well as isoprimeverose, xylosyl-cellobiose and the hepto-, octa- and nona saccharides of xyloglucan.

Gene structure determination

Genomic sequences of GH45 encoding genes were mined from publicly available draft genomes of *L. decemlineata* [27], *A. glabripennis* [25], *H. hampei* [28], *D. ponderosae* [24] and *S. oryzae* (unpublished; accession: SAMN08382431). The intron/exon structure was determined for each gene using splign [57], a spliced aligner.

Large phylogenetic analysis

We used the GH45 protein sequence from *Sitophilus oryzae* (ADU33247.1) as a BLASTp query against the NCBI's non-redundant protein library with an E-value threshold of $1E^{-3}$. We retrieved the 250 best blast hits (Additional file 1: Table S2), encompassing a majority of fungal sequences as well as various insect sequences (including Chrysomelidae, Curculionidae, Lamiinae and Collembola (=Entomobryomorpha)). Besides fungi and

insects, GH45 sequences from 10 nematodes, from one Tardigrade, one Rotifer and one bacterium, as well as a few uncharacterized protists from environmental samples, were among the 250 best BLAST hits. We complemented this dataset with predicted proteins from several Oribatid mites (10 sequences) and Collembola (4 sequences) retrieved from ncbi_tsa (Additional file 1: Table S2). This resulted in a collection of 264 sequences.

The set of 264 protein sequences was scanned against the Pfam v31 library of protein domains using the pfam_scan script with default parameters. All these proteins had one GH45 (Glyco_hydro_45 PF02015) domain on at least 98% of the expected lengths. Most of the fungal proteins possessed an ancillary carbohydrate-binding module (either CBM1 or CBM10), but this domain was not found in other species, except the bdelloid rotifer *Adineta ricciae.* We eliminated redundancy at 90% identity level between the 264 protein sequences; we used the CD-HIT Suite server [58] and reduced the dataset to 201 non-redundant sequences, while maintaining the diversity of clades.

The non-redundant sequences were aligned using MAFFT v7.271 [59] with the "—auto" option to allow us to automatically select the most appropriate alignment strategy (Additional file 2). We used trimal [60] to automatically discard columns that contained more than 50% of gaps in the alignment (–gt 0.5 option) and the maximum likelihood and the Bayesian methods to reconstruct phylogenetic trees. Maximum likelihood trees were reconstructed by RAxML version 8.2.9 [61] with an estimated gamma distribution of rates of evolution across sites and an automatic selection of the fittest evolutionary model (PROTGAMMAAUTO). Bootstrap replicates were automatically stopped upon convergence (–autoMRE). Bayesian trees were reconstructed by MrBayes version 3.2.6 [62] with an automatic estimation of the gamma distribution of rates of evolution across sites and a mixture of evolutionary models. The number of mcmc generations was stopped once the average deviation of split frequencies was below 0.05. Twenty-five percent of the trees were burnt for calculation of the consensus tree and statistics.

Amino acid alignment and Phytophaga-specific phylogenies

Sequences corresponding to Phytophaga GH45 proteins described in our previous studies were combined with those mined from several NCBI databases, such as the non-redundant protein database (ncbi_nr) and the transcriptome shotgun assembly database (ncbi_tsa) (Additional file 1: Table S3). In addition, transcriptome datasets generated from species of Phytophaga beetles were retrieved from the short-read archive (ncbi_sra) (Additional file 1: Table S3) and assembled using the CLC

workbench program version 11.0. Reads were loaded and quality trimmed before being assembled using standard parameters. The resulting assemblies were screened for contigs matching known beetle GH45 sequences through BLAST searches. The resulting contigs were then manually curated and used for further analysis. Amino acid alignments (Additional file 3) were carried out using MUSCLE version 3.7 implemented in MEGA7 (version 7.0.26) [63]. The maximum likelihood analysis was conducted in RAxML version 8.2.9. The best model of protein evolution was determined in MEGA7 using the 'find best DNA/protein models' tool. The best model was the 'Whelan and Goldman' (WAG) model, incorporating a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). The robustness of the analysis was tested using 1000 bootstrap replicates.

Additional files

Additional file 1: Table S1 to S4; Figs. S1 to S10 (legends included). (DOCX 22341 kb)

Additional file 2: Amino acid alignment used to build the Phylogeny presented in Fig. 3. (TXT 51 kb)

Additional file 3: Amino acid alignment of GH45 sequences derived from Phytophaga beetles used to build the Phylogeny presented in Fig. 4. (TXT 49 kb)

Availability of data and materials
All data generated or analyzed during this study are included in this published article [and its additional files].

Authors' contributions
AB, EGJD and YP conceived and designed the study; AB, EGJD and YP performed experiments; AB, EGJD and YP analyzed the data; AB, EGJD and YP wrote the paper. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate
Not applicable

Busch *et al. BMC Evolutionary Biology* (2019) 19:100

Page 13 of 14

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Entomology, Max Planck Institute for Chemical Ecology, Hans-Knoell-Str. 8, 07745 Jena, Germany. [2]INRA, Université Côte d'Azur, CNRS, ISA, Sophia Antipolis, France.

References
1. Chang MM, Chou T. Y. C., Tsao, G. T.: Structure, pretreatment and hydrolysis of cellulose In: Bioenergy Advances in Biochemical Engineering, vol 20. Springer, Berlin, Heidelberg; 1981. 15–42.
2. Ruel K, Nishiyama Y, Joseleau JP. Crystalline and amorphous cellulose in the secondary walls of *Arabidopsis*. Plant Sci. 2012;193-194:48–61.
3. Cosgrove DJ. Re-constructing our models of cellulose and primary cell wall assembly. Curr Opin Plant Biol. 2014;22:122–31.
4. Knox JP. Revealing the structural and functional diversity of plant cell walls. Curr Opin Plant Biol. 2008;11(3):308–13.
5. Saxena IM, Brown RM. A perspective on the assembly of cellulose-synthesizing complexes: possible role of korrigan and microtubules in cellulose synthesis in plants. In: Brown RM, Saxena IM, editors. Cellulose: molecular and structural biology. Dordrecht: Springer; 2007. p. 169–81.
6. Bayer EA, Chanzy H, Lamed R, Shoham Y. Cellulose, cellulases and cellulosomes. Curr Opin Struct Bio. 1998;8(5):548–57.
7. Chambost J.P. BMH, Cami B., Barras E. And Cattaneo J. : Erwinia cellulases. In: Civerolo E.L., Collmer A., Davis R.E., Gillaspie A.G. (eds) Plant pathogenic Bacteria. Current plant science and biotechnology in agriculture, vol 4. Springer, Dordrecht; 1987: 150–159.
8. Py B, Bortoli-German I, Haiech J, Chippaux M, Barras F. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His98 and Glu133 residues for catalysis. Protein Eng. 1991;4(3):325–33.
9. Schulein M. Enzymatic properties of cellulases from *Humicola insolens*. J Biotechnol. 1997;57(1–3):71–81.
10. Breznak JA, Brune A. Role of microorganisms in the digestion of lignocellulose by termites. Annu Rev Entomol. 1994;39:453–87.
11. Rincon MT, McCrae SI, Kirby J, Scott KP, Flint HJ. EndB, a multidomain family 44 cellulase from *Ruminococcus flavefaciens* 17, binds to cellulose via a novel cellulose-binding module and to another *R. flavefaciens* protein via a dockerin domain. Appl Environ Microbiol. 2001;67(10):4426–31.
12. Smant G, Stokkermans JP, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Henrissat B, Davis EL, et al. 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. 1998;95(9):4906–11.
13. Watanabe H, Noda H, Tokuda G, Lo N. A cellulase gene of termite origin. Nature. 1998;394(6691):330–1.
14. Girard C, Jouanin L. Molecular cloning of cDNAs encoding a range of digestive enzymes from a phytophagous beetle, *Phaedon cochleariae*. Insect Biochem Mol Biol. 1999;29(12):1129–42.
15. Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*. FEBS Lett. 2004;572(1–3):201–5.
16. Sakamoto K, Toyohara H. Molecular cloning of glycoside hydrolase family 45 cellulase genes from brackish water clam *Corbicula japonica*. Comp Biochem Physiol B Biochem Mol Biol. 2009;152(4):390–6.
17. Faddeeva-Vakhrusheva A, Derks MF, Anvar SY, Agamennone V, Suring W, Smit S, van Straalen NM, Roelofs D. Gene family evolution reflects adaptation to soil environmental stressors in the genome of the collembolan *Orchesella cincta*. Genome Biol Evol. 2016;8(7):2106–17.
18. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH. Diversity of beetle genes encoding novel plant cell wall degrading enzymes. PLoS One. 2010;5(12):e15635.
19. Takahashi M, Takahashi H, Nakano Y, Konishi T, Terauchi R, Takeda T. Characterization of a cellobiohydrolase (MoCel6A) produced by *Magnaporthe oryzae*. Appl Environ Microbiol. 2010;76(19):6583–90.
20. Kostylev M, Wilson D. Synergistic interactions in cellulose hydrolysis. Biofuels. 2012;3(1):61–70.
21. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:490–5.
22. Aspeborg H, Coutinho PM, Wang Y, Brumer H 3rd, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol Biol. 2012;12:186.
23. Watanabe H, Tokuda G. Cellulolytic systems in insects. Annu Rev Entomol. 2010;55:609–32.
24. Keeling CI, Yuen MM, Liao NY, Docking TR, Chan SK, Taylor GA, Palmquist DL, Jackman SD, Nguyen A, Li M, et al. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 2013;14(3):R27.
25. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn SJ, Arsala D, et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biol. 2016;17(1):227.
26. Pauchet Y, Saski CA, Feltus FA, Luyten I, Quesneville H, Heckel DG. Studying the organization of genes encoding plant cell wall degrading enzymes in *Chrysomela tremula* provides insights into a leaf beetle genome. Insect Mol Biol. 2014;23(3):286–300.
27. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, Brevik K, Cappelle K, Chen MM, Childers AK, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Sci Rep. 2018;8(1):1931.
28. Vega FE, Brown SM, Chen H, Shen E, Nair MB, Ceja-Navarro JA, Brodie EL, Infante F, Dowd PF, Pain A. Draft genome of the most devastating insect pest of coffee worldwide: the coffee berry borer, *Hypothenemus hampei*. Sci Rep. 2015;5:12525.
29. Marvaldi AE, Duckett CN, Kjer KM, Gillespie JJ. Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). Zool Scr. 2009;38(1):63–77.
30. Kirsch R, Wielsch N, Vogel H, Svatos A, Heckel DG, Pauchet Y. Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle. BMC Genomics. 2012;13:587.
31. Pauchet Y, Kirsch R, Giraud S, Vogel H, Heckel DG. Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle *Apriona japonica*. Insect Biochem Mol Biol. 2014;49:1–13.
32. Lee SJ, Kim SR, Yoon HJ, Kim I, Lee KS, Je YH, Lee SM, Seo SJ, Dae Sohn H, Jin BR. cDNA cloning, expression, and enzymatic activity of a cellulase from the mulberry longicorn beetle, *Apriona germari*. Comp Biochem Physiol B Biochem Mol Biol. 2004;139(1):107–16.
33. Mei HZ, Xia DG, Zhao QL, Zhang GZ, Qiu ZY, Qian P, Lu C. Molecular cloning, expression, purification and characterization of a novel cellulase gene (Bh-EGasel) in the beetle *Batocera horsfieldi*. Gene. 2015;576(1):45–51.
34. Chang CJ, Wu CP, Lu SC, Chao AL, Ho TH, Yu SM, Chao YC. A novel exo-cellulase from white spotted longhorn beetle (*Anoplophora malasiaca*). Insect Biochem Mol Biol. 2012;42(9):629–36.
35. Valencia A, Alves AP, Siegfried BD. Molecular cloning and functional characterization of an endogenous endoglucanase belonging to GHF45 from the western corn rootworm, *Diabrotica virgifera virgifera*. Gene. 2013;513(2):260–7.
36. Busch A, Kunert G, Wielsch N, Pauchet Y. Cellulose degradation in *Gastrophysa viridula* (Coleoptera: Chrysomelidae): functional characterization of two CAZymes belonging to glycoside hydrolase family 45 reveals a novel enzymatic activity. Insect Mol Biol. 2018;27(5):633–50.
37. Davies GJ, Tolley SP, Henrissat B, Hjort C, Schulein M. Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 a resolution. Biochemistry. 1995;34(49):16210–20.
38. DeBoy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, Mohamoud Y, Watkins K, Henrissat B, Gilbert HJ, et al. Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. J Bacteriol. 2008;190(15):5455–63.
39. Sheppard PO, Grant FJ, Oort PJ, Sprecher CA, Foster DC, Hagen FS, Upshall A, McKnight GL, O'Hara PJ. The use of conserved cellulase family-specific

sequences to clone cellulase homologue cDNAs from *Fusarium oxysporum*. Gene. 1994;150(1):163–7.

40. O'Connor RM, Fung JM, Sharp KH, Benner JS, McClung C, Cushing S, Lamkin ER, Fomenkov AI, Henrissat B, Londer YY, et al. Gill bacteria enable a novel digestive strategy in a wood-feeding mollusk. Proc Natl Acad Sci U S A. 2014;111(47):E5096–104.

41. Xu BZ, Hagglund P, Stalbrand H, Janson JC. Endo-beta-1,4-mannanases from blue mussel, *Mytilus edulis*: purification, characterization, and mode of action. J Biotechnol. 2002;92(3):267–77.

42. Rahman MM, Inoue A, Ojima T. Characterization of a GHF45 cellulase, AkEG21, from the common sea hare *Aplysia kurodai*. Front Chem. 2014;2(60).

43. Palomares-Rius JE, Hirooka Y, Tsai IJ, Masuya H, Hino A, Kanzaki N, Jones JT, Kikuchi T. Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. BMC Evol Biol. 2014;14:69.

44. Song JM, Hong SK, An YJ, Kang MH, Hong KH, Lee YH, Cha SS. Genetic and structural characterization of a thermo-tolerant, cold-active, and acidic endo-beta-1,4-glucanase from Antarctic springtail, *Cryptopygus antarcticus*. J Agric Food Chem. 2017;65(8):1630–40.

45. Calderon-Cortes N, Watanabe H, Cano-Camacho H, Zavala-Paramo G, Quesada M. cDNA cloning, homology modelling and evolutionary insights into novel endogenous cellulases of the borer beetle *Oncideres albomarginata chamela* (Cerambycidae). Insect Mol Biol. 2010;19(3):323–36.

46. Eyun SI, Wang H, Pauchet Y, Ffrench-Constant RH, Benson AK, Valencia-Jimenez A, Moriyama EN, Siegfried BD. Molecular evolution of glycoside hydrolase genes in the western corn rootworm (*Diabrotica virgifera virgifera*). PLoS One. 2014;9(4):e94052.

47. Lee SJ, Lee KS, Kim SR, Gui ZZ, Kim YS, Yoon HJ, Kim I, Kang PD, Sohn HD, Jin BR. A novel cellulase gene from the mulberry longicorn beetle, *Apriona germari*: gene structure, expression, and enzymatic activity. Comp Biochem Physiol B Biochem Mol Biol. 2005;140(4):551–60.

48. Szydlowski L, Boschetti C, Crisp A, Barbosa EG, Tunnacliffe A. Multiple horizontally acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the bdelloid rotifer *Adineta ricciae*. Gene. 2015;566(2):125–37.

49. Mcgavin M, Forsberg CW. Isolation and characterization of endoglucanase-1 and endoglucanase-2 from *Bacteroides succinogenes* S85. J Bacteriol. 1988;170(7):2914–22.

50. Busch A, Kunert G, Heckel DG, Pauchet Y. Evolution and functional characterization of CAZymes belonging to subfamily 10 of glycoside hydrolase family 5 (GH5_10) in two species of phytophagous beetles. PLoS One. 2017;12(8):e0184305.

51. Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee SJ, Yeats TH, Egan AN, Doyle JJ, et al. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proc Natl Acad Sci U S A. 2012;109(11):4197–202.

52. Scheller HV, Ulvskov P. Hemicelluloses. Annu Rev Plant Biol. 2010;61:263–89.

53. Pauly M, Gille S, Liu L, Mansoori N, de Souza A, Schultink A, Xiong G. Hemicellulose biosynthesis. Planta. 2013;238(4):627–42.

54. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008;9(12):938–50.

55. Kirsch R, Gramzow L, Theissen G, Siegfried BD, Ffrench-Constant RH, Heckel DG, Pauchet Y. Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: key events in the evolution of herbivory in beetles. Insect Biochem Mol Biol. 2014;52:33–50.

56. Zhang YH, Cui J, Lynd LR, Kuang LR. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. Biomacromolecules. 2006;7(2):644–8.

57. Kapustin Y, Souvorov A, Tatusova T, Lipman D. Splign: algorithms for computing spliced alignments with identification of paralogs. Biol Direct. 2008;3:20.

58. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.

59. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010;26(5):680–2.

60. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.

61. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

62. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

63. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.