Analyzing the Substrate Specificity of a Class of Long-Horned-Beetle-Derived Xylanases by Using Synthetic Arabinoxylan Oligo- and Polysaccharides

Yannick Pauchet,*[a] Colin Ruprecht,[b] and Fabian Pfrengle[b, c]

Xylophagous long-horned beetles thrive in challenging environments. To access nutrients, they secrete plant-cell-wall-degrading enzymes in their gut fluid; among them are cellulases of the subfamily 2 of glycoside hydrolase family 5 (GH5_2). Recently, we discovered that several beetle-derived GH5_2s use xylan as a substrate instead of cellulose, which is unusual for this family of enzymes. Here, we analyze the substrate specificity of a GH5_2 xylanase from the beetle *Apriona japonica* (AJAGH5_2-1) using commercially available substrates and synthetic arabinoxylan oligo- and polysaccharides. We demonstrate that AJAGH5_2-1 processes arabinoxylan polysaccharides in a manner distinct from classical xylanase families such as GH10 and GH11. AJAGH5_2-1 is active on long oligosaccharides and cleaves at the non-reducing end of a substituted xylose residue (position +1) only if: 1) three xylose residues are present upstream and downstream of the cleavage site, and 2) xylose residues at positions −1, −2, +2 and +3 are not substituted.

Introduction

Insects are more numerous than any other kind of animal, and the Coleoptera (beetle family) is the group of insects with the highest number of known species; as such, they have colonized almost every ecological niche on earth. With around 130 000 species distributed worldwide—fifty percent of the plant-feeding insect diversity—the monophyletic Phytophaga clade of beetles represents the largest group of herbivorous insects on the planet.[1] Phytophaga beetles encompass the superfamilies Chrysomeloidea (leaf beetles and long-horned beetles) and Curculionoidea (weevils and bark beetles). Comprising some of the most efficient herbivorous insects, many of these are important pests of agriculture and forestry, namely, the Colorado potato beetle *Leptinotarsa decemlineata* (Chrysomelidae), the western corn rootworm *Diabrotica virgifera virgifera* (Chrysomelidae), the Asian long-horned beetle *Anoplophora glabripennis* (Cerambycidae) and the mountain pine beetle *Dendroctonus ponderosae* (Curculionidae). The causes of their extraordinarily successful radiation are widely debated, but the evolution of specialized trophic interactions with plants is assumed to have played an important role in this process.[1c,2]

The ability of Phytophaga beetles to deal with the structural polysaccharides of the plant cell wall, which make up the bulk of their food, is essentially due to the presence of so-called plant-cell-wall-degrading enzymes (PCWDEs) in their gut fluids.[3] Targeting cellulose, hemicelluloses and pectins, these carbohydrate-active enzymes (CAZymes) are encoded by several medium-sized gene families endogenous to the beetle's genome.[4] Gene families encoding putative cellulolytic glycoside hydrolase (GH) family 9 (GH9), family 45 (GH45), family 48 (GH48) and the pectolytic family 28 (GH28) were found to be broadly distributed in Phytophaga beetles[3b,4d,5] In contrast, other families of PCWDEs are more restricted to specific groups of beetles, such as the putative cellulolytic subfamily 2 of GH5 (GH5_2) for which corresponding genes have to date been identified, in insects, only in species of long-horned beetles.[4b,d,6]

The beetle family Cerambycidae Latreille (long-horned beetles)—with an estimated number of species above 35 000—contains the most diverse radiation of wood-feeding animals on earth.[2b,7] Cerambycid larvae bore deep into healthy, moribund, recently dead or decomposing woody plants. Xylophagous insects, such as long-horned beetles, have evolved to thrive in highly challenging habitats; to efficiently access the nutrients present in such sub-optimal environments, they have to cope with the lignocellulose barrier. The presence of PCWDEs in the gut fluid of wood-boring larvae of the Cerambycidae has long been known.[8] Initially these enzymes were thought to be secreted by symbiotic yeasts located within the insect's gut either in epithelial cells or in specific tissue called

[a] Dr. Y. Pauchet
Department of Entomology, Max Planck Institute for Chemical Ecology
Hans-Knoell-Strasse 8, 07745 Jena (Germany)
E-mail: ypauchet@ice.mpg.de

[b] Dr. C. Ruprecht, Dr. F. Pfrengle
Department of Biomolecular Systems
Max Planck Institute of Colloids and Interfaces
Am Mühlenberg 1, 14476 Potsdam (Germany)

[c] Dr. F. Pfrengle
Institute of Chemistry and Biochemistry, Freie Universität Berlin
Takustrasse 3, 14195 Berlin (Germany)

Supporting information and the ORCID identification numbers for the authors of this article can be found under https://doi.org/10.1002/cbic.201900687.

the mycetome. However, this hypothesis was eventually ruled out by several studies from the 1930s, reviewed in,[8b] which showed that 1) symbiont-free species of cerambycid could break down cellulose as efficiently as species harboring those yeast symbionts; and 2) these symbiotic yeasts, when isolated and cultured, were not cellulolytic. Thus, symbiotic-independent cellulose and hemicellulose digestion is proposed to occur in cerambycid beetle larvae.

In the 1970s and 80s, PCWDEs were purified and characterized from several species of cerambycid beetles. In general, the gut of cerambycid larvae is characterized by a high pectolytic activity as well as activity against hemicellulose polysaccharides, such as xylan, β-1,3-glucans and mannans. Activity against amorphous cellulose is always detected, whereas activity against crystalline cellulose is either low or completely absent, except in species feeding on decaying or rotten wood. In the latter case, the enzymes active on crystalline cellulose seem to be provided by the fungi ingested by the larvae while feeding on decaying/rotten wood.[3c,9] The genes encoding PCWDEs in long-horned beetles have long remained elusive and are still largely unknown. The first cellulase gene derived from the genome of a long-horned beetle, reported in 2003 from the mulberry long-horned beetle *Psacothea hilaris*, encodes a GH5_2 *endo*-β-1,4-glucanase.[6e] Since then, several GH5_2 cellulases have been cloned from other species of cerambycid beetles.[4b,6b,d,f]

According to the CAZy database,[10] to date 120 GH5_2 proteins, most of which are derived from bacteria, have been functionally characterized, and almost all are cellulolytic (*endo*-β-1,4-glucanases; EC 3.2.1.4), whereas a few bacterial enzymes specifically hydrolyze chitosan (chitosan *N*-acetylglucosaminohydrolase, EC 3.2.1.132). Similarly, most of the GH5_2 derived from long-horned beetles which have been functionally characterized are cellulases (EC3.2.1.4).[6b,e,f] However, we recently discovered that some GH5_2 proteins from two species of long-horned beetles are able to hydrolyze xylan instead of cellulose, making them *endo*-β-1,4-xylanases (EC 3.2.1.8); such an ability is uncommon in this family of proteins.[4b,6d] In fact, GH5 xylanases do exist but are restricted to subfamilies 4 (GH5_4) and 21 (GH5_21).[11] Given that most larvae of long-horned beetles are xylophagous[7] and that xylan is the most abundant hemicellulose polysaccharide in wood material, possessing the ability to break down xylan is clearly advantageous for these insects.

To better understand how this novel class of xylanases work, we tested the catalytic abilities of a GH5_2 protein (*AJA*GH5_2) derived from the long-horned beetle *Apriona japonica*[6d] against a variety of substrates. First, by using commercially available poly- and oligosaccharides, we confirm that *AJA*GH5_2 is a specific *endo*-β-1,4-xylanase and is unable to break down cellulose or other plant-cell-wall-associated polysaccharides. We also show that *AJA*GH5_2 possesses the ability to use glucurono- and arabinoxylan polymers harboring various degree of substitution as substrates. Second, to obtain more details on how *AJA*GH5_2 breaks down xylan, we used a combination of synthetic arabinoxylan oligomers generated by automated glycan assembly[12] and artificial polysaccharides synthesized by the enzymatic polymerization of arabinoxylan oligosaccharide fluorides.[13] We demonstrate that *AJA*GH5_2 is the first member of a novel class of *endo*-β-1,4-xylanases with a distinct strategy for cleaving arabinoxylan substrates compared to classical xylanases from the GH10 and GH11 families. We show that *AJA*GH5_2 can only hydrolyze long oligosaccharides and can cleave at the non-reducing end of a substituted xylose residue (position +1) only if 1) three xylose residues are present upstream and downstream of the cleavage site and if 2) xylose residues at positions −1, −2, +2 and +3 are not substituted.

Results

AJAGH5_2-1 is an *endo*-acting glycoside hydrolase specifically active on xylan polymers and oligomers

*AJA*GH5_2-1 was expressed in *Sf*9 insect cells as described earlier,[6d] and we performed simple end-point measurements by incubating, overnight at 40 °C, the crude enzyme extract with a series of polysaccharides usually associated with the plant cell wall, namely carboxymethylcellulose (CMC) and regenerated amorphous cellulose, as well as the hemicelluloses xyloglucan, gluco- and galactomannan, and beechwood xylan (a form of glucuronoxylan; Figure 1 A). We analyzed the resulting breakdown products by thin-layer chromatography (TLC). Only the incubation of *AJA*GH5_2-1 with beechwood xylan resulted in the appearance of breakdown products (Figure 1 A). The breakdown products which accumulated were mostly xylobiose, xylotriose and xylotetraose, as well as long oligomers. In addition, the presence of double signals close to the xylotriose and xylotetraose standards may indicate the presence of glucuronoxylotriose as well as glucuronoxylotetraose (Figure 1 A).

We subsequently incubated *AJA*GH5_2-1, under the same conditions as above, with xylan polymers harboring various levels of substitutions. As substrates, we chose beechwood xylan (a glucuronoxylan; substitution: GlcOMe/Xyl = 1:10), arabinoxylan from rye (substitution: Ara/Xyl = 38:62) and arabinoxylan from wheat (substitution: Ara/Xyl = 22:78). Breakdown products visible with TLC demonstrated that *AJA*GH5_2-1 was able to take apart these three xylan polymers (Figure 1 B). As the TLC revealed, the degree of substitution has an influence on the activity of the enzyme: the more substituted the xylan polymer is the less abundant small oligomers, such as xylobiose, xylotriose and xylotetraose, are (Figure 1 B).

We further asked what minimum size a xylan oligomer must be to be hydrolyzed by *AJA*GH5_2-1. To find out, we incubated *AJA*GH5_2-1 with xylan oligomers, ranging from xylobiose to xylohexaose, and resolved the resulting end-products by TLC (Figure 1 B). Whereas breakdown products were present when *AJA*GH5_2-1 was incubated with xylohexaose, none was observed when xylopentaose was used as a substrate, indicating that the hexamer is likely the shortest xylan oligomer which can be hydrolyzed by *AJA*GH5_2-1. According to the pattern of breakdown products obtained, *AJA*GH5_2-1 hydrolyzed xylohexaose either in two molecules of xylotriose or in one molecule of xylotetraose plus one molecule of xylobiose (Figure 1 B). No breakdown products resulted from the incubation

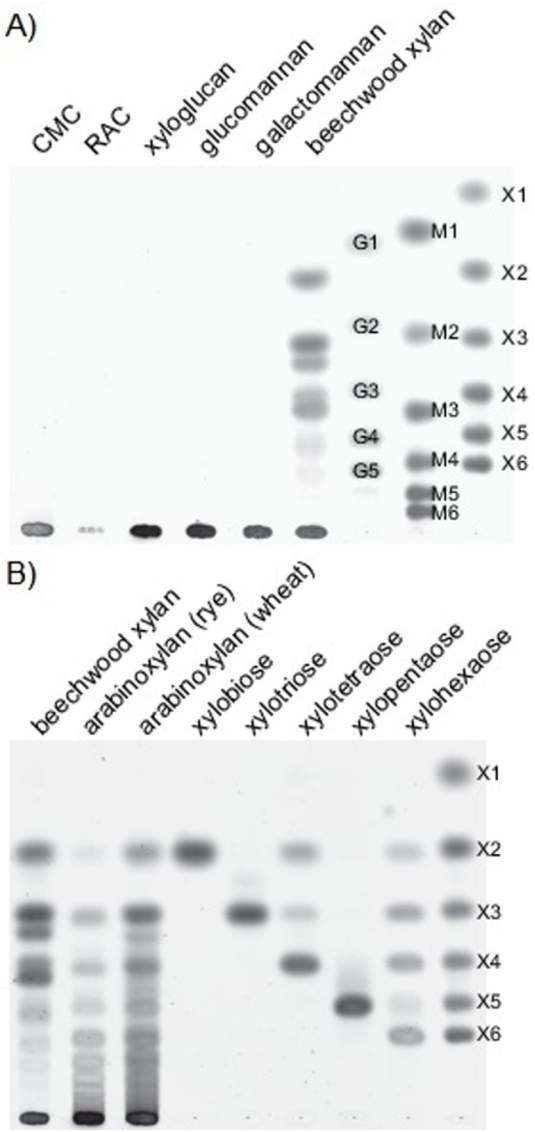

Figure 1. *A. japonica* GH5_2-1 is a specific *endo*-β-1,4-xylanase. A) Thin-layer chromatography (TLC) analyses of the hydrolysis end-products of polysaccharides typically found in the plant cell wall by *A. japonica* GH5_2-1. Crude enzyme extracts, derived from the culture media of transfected Sf9 cells, were incubated with carboxymethylcellulose (CMC), regenerated amorphous cellulose (RAC), glucomannan, galactomannan and beechwood xylan in McIlvaine buffer pH 5.0 at 40 °C for 16 h before being spotted on TLC plates. The standards used for the cellulose-derived substrates were glucose (G1) to cellopentaose (G5). The standards used for the mannan-derived substrates were mannose (M1) to mannohexaose (M6). The standards used for the xylan-derived substrates were xylose (X1) to xylohexaose (X6). B) TLC analyses of the hydrolysis end-products of various xylan poly- and oligosaccharides by *A. japonica* GH5_2-1. Crude enzyme extracts were incubated as described above with beechwood xylan (a glucuronoxylan; substitution: GlcOMe/Xyl = 1:10), arabinoxylan from rye (substitution: Ara/Xyl = 38:62), arabinoxylan from wheat (substitution: Ara/Xyl = 22:78), as well as, xylan oligomers ranging from xylobiose to xylohexaose. The standards used are the same as described above.

of *AJA*GH5_2-1 with xylobiose and xylotriose (Figure 1 B). Unexpectedly, the incubation of *AJA*GH5_2-1 with xylotetraose resulted in the appearance of xylobiose and a minute amount of xylotriose (Figure 1 B). Although the presence of xylobiose would be logical and would indicate that *AJA*GH5_2-1 could hydrolyze xylotetraose in two xylobiose molecules, having xylotriose without xylose in this context does not make much sense. In addition, the apparent inability of *AJA*GH5_2-1 to hydrolyze xylopentaose renders the hydrolysis pattern of xylotetraose even more dubious. To us, the only explanation for the presence of xylobiose and xylotriose as a result of the incubation of *AJA*GH5_2-1 with xylotetraose is the potential ability of *AJA*GH5_2-1 to perform a transglycosylation reaction in the presence of an excess amount of xylotetraose. Currently this assumption remains a possibility which must be tested in future analyses. Yet clearly *AJA*GH5_2-1 possesses the ability to break down xylan polymers, even highly substituted polymers, as well as unsubstituted oligomers the size of xylohexaose or longer.

The activity of AJAGH5_2-1 is dependent on the size of xylan oligomers and the location of substitutions

To address how arabinose substitutions of the xylan chain influence the activity of *AJA*GH5_2-1, we incubated the enzyme with a series of synthetic arabinoxylan oligomers generated by automated glycan assembly.[12] The resulting breakdown products were analyzed by HPLC coupled to a mass spectrometer and an evaporative light-scattering detector (ELSD). First, we used linear xylan oligosaccharides, namely xylotetraose, xylohexaose and xylooctaose, which we incubated for 16 h at 40 °C with *AJA*GH5_2-1 (Figure 2). Under these assay conditions, no breakdown products were produced by the action of *AJA*GH5_2-1 on the tetrasaccharide. These results were in contrast to those obtained when we incubated *AJA*GH5_2-1 with a commercially available xylotetraose (Figure 1 B). The xylohexaose produced by automated glycan assembly was completely hydrolyzed by *AJA*GH5_2-1 in two xylotriose molecules, one with a free reducing end and the other with the aminopentyl linker at its reducing end (Figure 2). Again, this result was in contrast with what we observed using a commercially available xylohexaose as a substrate for which xylotriose as well as xylotetraose and xylobiose were visible using TLC (Figure 1 B). The octasaccharide was also fully hydrolyzed in xylobiose, xylobiose with the aminopentyl linker at its reducing end, xylotriose and xylotriose with the aminopentyl linker at its reducing end (Figure 2).

To understand how arabinose substitutions of xylan affect the activity of *AJA*GH5_2-1, we incubated the enzyme with a series of xylohexaose molecules; these were produced by automated glycan assembly and harbored one or two arabinose substitutions on carbon 2 or carbon 3 of a given xylose residue at various positions of a xylohexaose backbone (Figure 3). *AJA*GH5_2-1 could not use a xylohexaose molecule harboring an arabinose substitution on either carbon 2 or carbon 3 of the third xylose residue starting from the non-reducing end as a substrate (Figure 3). Similarly, the presence of two arabinose substitutions—one on the second and one on the penultimate xylose residue starting from the non-reducing end—impaired the ability of *AJA*GH5_2-1 to use this xylohexaose as a substrate. In contrast, *AJA*GH5_2-1 could hydrolyze a xylohexaose

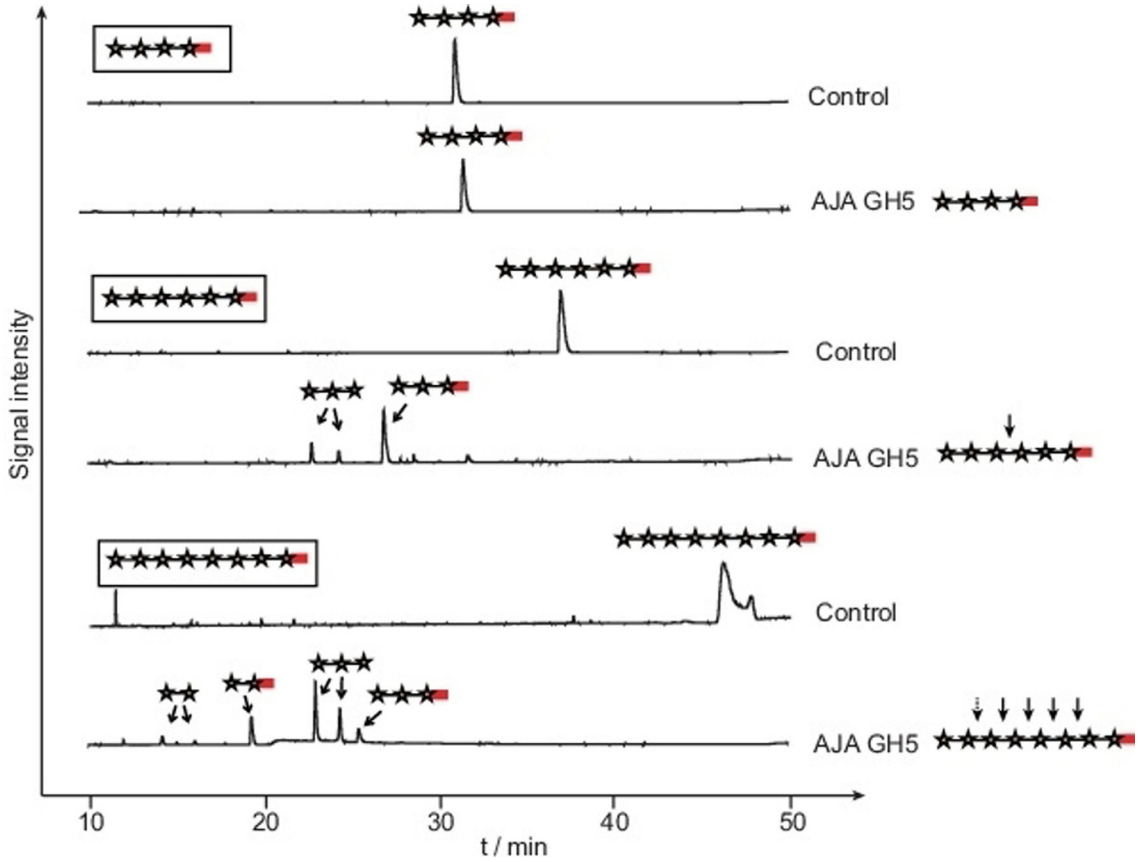

Figure 2. Enzymatic digestion products of linear xylan oligosaccharides by *A. japonica* GH5_2 xylanase. HPLC analysis of reactions products was performed using a porous graphitized carbon (PGC) column and detected with an evaporative light-scattering detector (ELSD) and mass spectrometry (MS). Shown are the respective ELSD traces. Products were identified using their corresponding retention time and mass. Substrates are indicated in boxes. The red bars denote an aminopentyl linker at the reducing end of the oligosaccharides. Structures without this red bar have a free reducing end. For the breakdown products with free reducing ends, the corresponding alpha and beta anomers could be separated on the HPLC, resulting in two peaks for these compounds. The cutting sites of the xylanase are indicated on the right with arrows. Sites that are not preferred by the enzyme are denoted with dashed arrows.

harboring an arabinose substitution located on carbon 2 of the first xylose residue at the nonreducing end (Figure 3). Altogether, these results showed that the presence and position of arabinose substitutions on the xylan chain strongly influence the catalytic abilities of *AJA*GH5_2-1. Moreover, the fact that only a single of these arabinosylated oligosaccharides could be cleaved by *AJA*GH5_2-1 confirmed that this enzyme requires oligosaccharides with at least three xylose residues upstream and downstream of the cleavage site to be active.

To further investigate how the catalytic activity of *AJA*GH5_2-1 is influenced by arabinose substitutions, we used arabinoxylan polysaccharides synthesized by the enzymatic polymerization of arabinoxylan oligosaccharide fluorides.[13] These artificial xylan polysaccharides are furnished with a perfectly regular pattern of arabinose substitution. We tested three types of polysaccharides, each of which was made of a different repeated motif (Figure 4). The first is a repetition of a trimer being substituted by an arabinose on carbon 3 of the first xylose starting from the non-reducing end. The second polysaccharide is similar to the first except that the decoration is located on carbon 2 of the first xylose residue starting from the non-reducing end. The third polysaccharide, a repetition of a tetramer, is substituted by an arabinose on carbon 3 of the first xylose starting from the non-reducing end. *AJA*GH5_2-1 could hydrolyze all three types of polysaccharide in a similar way at the non-reducing end of a substituted xylose residue (Figure 4).

In summary, the use of various synthetic substrates allowed us to deduce the general requirements for arabinose substitutions and the minimum length of xylan oligosaccharides which are tolerated by the *A. japonica* GH5_2 *endo*-β-1,4-xylanase (Figure 5). *AJA*GH5_2-1 can accommodate an arabinose substitution at the +1 site only if a minimum of three xylose molecules are present downstream of the cleavage site, with the +2 and +3 sites being unsubstituted. In addition, a minimum of three xylose residues must be located upstream of the cleavage site (−1 to −3 positions) for the enzyme to cleave. An arabinose substitution at the −3 position can be tolerated by the enzyme. Whether the arabinose substitution occurs at carbon 2 or carbon 3 of the xylose at the −3 and/or at the +1 site does not seem to influence the activity of *AJA*GH5_2-1 (Figure 5).

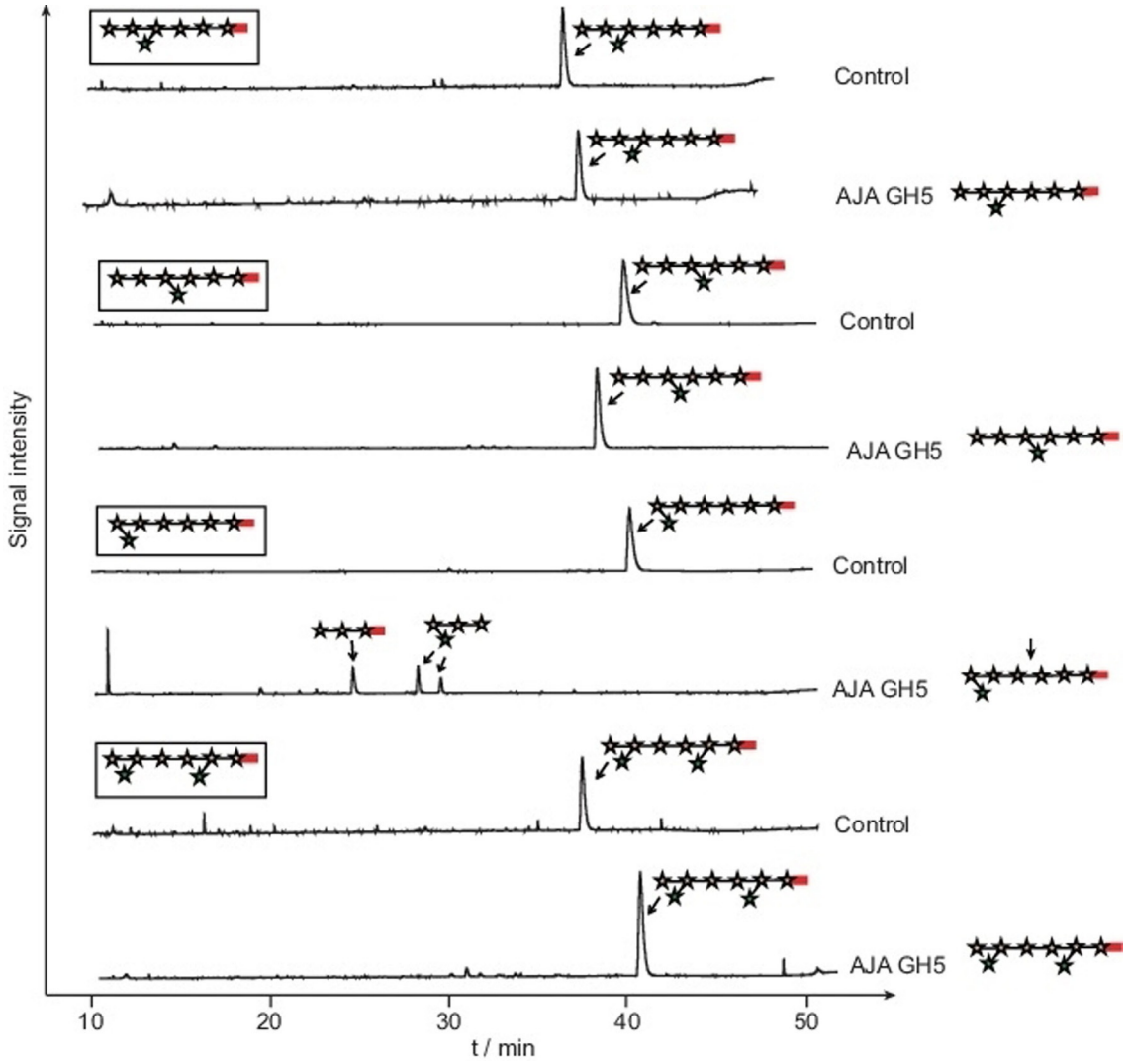

Figure 3. Enzymatic digestion products of arabinosylated xylan oligosaccharides by *A. japonica* GH5_2 xylanase. Substrates are indicated in boxes. The red bars denote an aminopentyl linker at the reducing end of the oligosaccharides. Structures without this red bar have a free reducing end. Arabinose substitutions (green stars) on carbon 2 or carbon 3 of a xylose residue (brown star) are depicted by lower right or lower left connections to the xylose, respectively. The corresponding alpha and beta anomers of the arabinosylated xylotriose could be separated by HPLC, resulting in two peaks for each of these compounds. The cutting sites of the xylanase are indicated on the right with arrows. Note that retention times using PGC-HPLC may vary between control and treatment. MS data were used to annotate the peaks in the shown ELSD traces.

GH5_2 xylanases might be distributed in other species of long-horned beetles and are not restricted to *A. japonica*

To address whether the presence of a gene encoding a GH5_2 xylanase is unique to *A. japonica* or whether the genomes of other long-horned beetles also harbor these, we screened public databases for GH5_2 sequences derived from other long-horned beetles and reconstructed their phylogenetic relationships by maximum likelihood (Figure 6). We recovered 28 GH5_2 sequences from eight species of long-horned beetles (Table S1), all belonging to the subfamily Lamiinae. Some of these GH5_2 proteins have been functionally characterized, most as cellulases. Qualitative enzyme assays have shown that another GH5_2 protein (AGL1), encoded by the genome of the Asian long-horned beetle *A. glabripennis*, is active on beechwood xylan.[4b] We used these 28 sequences to perform a maximum likelihood-inferred phylogeny (Figure 6). The cerambycid-derived GH5_2 sequences formed six highly supported clades of orthologous sequences (clades I to VI; Figure 6). *AJA*GH5_2-1 (AJA1) and AGL1 are orthologs and are located in clade I with sequences from two other long-horned beetle species, suggesting that GH5_2 xylanases may be distributed in other species of long-horned beetles. Of course, cloning and functionally characterizing GH5_2 proteins from clade I will be necessary to make sure that this clade is xylanase-specific.

Discussion

Here, we demonstrate that *AJA*GH5_2-1 is indeed a new type of xylanase with specific characteristics that differ from those of typical GH10 or GH11 xylanases.[12,14] Whereas GH10 and GH11 xylanases can usually break down relatively small arabi-

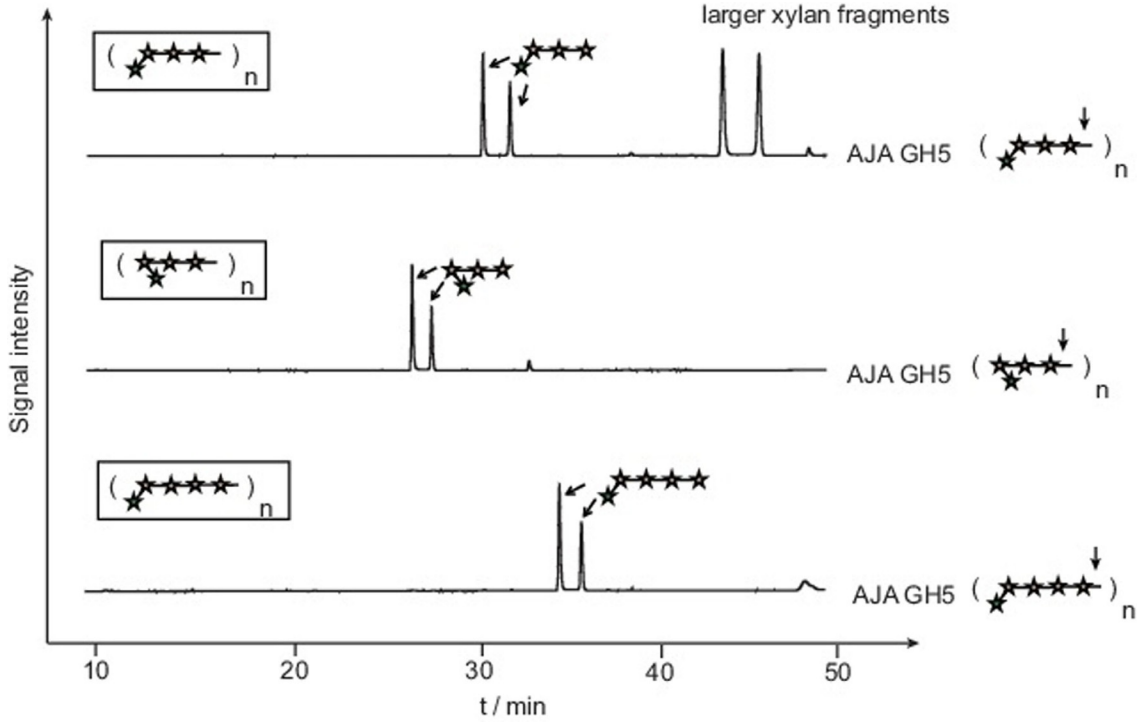

Figure 4. Enzymatic digestion products of synthetic xylan polysaccharides by *A. japonica* GH5_2 xylanase. Substrates are indicated in boxes, with the repeating unit of the polysaccharide in brackets. The larger xylan fragments in the upper panel could be digested with a commercially available GH10 xylanase (Megazyme; accession number: P14768.2). More detailed information on their structure could not be obtained. The cutting sites of the xylanase are indicated on the right with arrows.

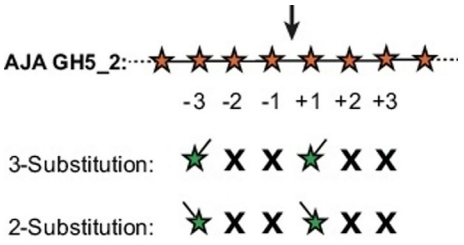

Figure 5. Results of active site mapping of the *A. japonica* GH5_2 xylanase. The cutting site of the xylanase is indicated with an arrow. The tolerated arabinose substitutions at C2 or C3 positions of the xyloses adjacent to the cutting site are indicated below. Positions that should not be substituted are marked with an "x". Note that, in general, substrates with at least three xyloses on each side of the cutting site are required for the activity of *A. japonica* GH5_2 xylanase, as only one of the arabinosylated oligosaccharides could be digested.

noxylan oligosaccharides,[12] *AJA*GH5_2-1 requires long oligosaccharides—at least hexamers—to be active. GH5 xylanases have been described earlier, but are distributed in subfamilies 4 and 21 (GH5_4 and GH5_21).[11] Most GH5_2 proteins that have already been characterized are *endo*-β-1,4-glucanases acting on amorphous cellulose, a linear polysaccharide that harbors no substitutions at all.[11] In contrast, xylans are substituted in nature, either by glucuronic acid or by arabinose linked to xylose through either α-1,2 or α-1,3 glycosidic bonds. We show that the active site of *AJA*GH5_2-1 can accommodate substitutions and can cleave at the non-reducing end of a substituted xylose only if at least three xylose residues are present upstream and downstream of the cleavage site. To understand the changes necessary for the active site of *AJA*GH5_2-1 to accommodate a substituted substrate compared to a classical GH5_2 cellulase would require to resolve the structure of this protein and to compare it with the one of a GH5_2 cellulase. Interestingly, each genome/transcriptome of cerambycid beetles which have been investigated so far encodes at least one GH5_2 cellulase.[4b,6d–f] The corresponding proteins are orthologous and cluster together in clade V in the phylogeny presented in Figure 6. Resolving the structure of such a cerambycid-derived GH5_2 cellulase in order to make a direct comparison with the structure of *AJA*GH5_2-1 will become our future priority.

Here we observed discrepancies between the results we obtained using commercial xylan oligosaccharides (Figure 1B) and those obtained using oligosaccharides produced by automated glycan assembly (Figure 2). For example, we obtained not only xylotriose from the action of *AJA*GH5_2-1 on the commercially available xylohexaose (Figure 1B), but also xylotetraose and xylobiose, whereas we obtained only xylotriose using the xylohexaose generated by automated glycan assembly (Figure 2). These differences cannot be attributed to differences in assay conditions between the two types of experiments. In fact, the same batch of crude enzyme extract was used in both types of experiments and the concentrations of oligosaccharides used were very similar. We used slightly more enzyme extract in the assays with commercially available substrates relative to those performed with the synthetic substrates (14 vs. 10 μL). We cannot exclude the possibility that this difference

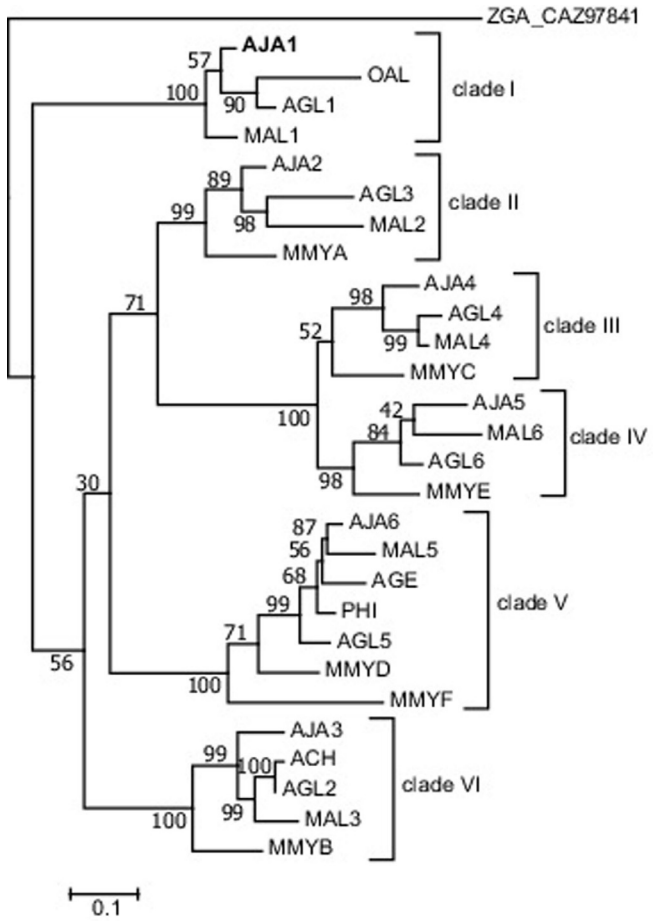

Figure 6. Phylogenetic relationships between GH5_2 family members of various Cerambycid beetles. A maximum likelihood analysis was performed with 1000 bootstrap replicates using the program IQTREE. The best model of protein evolution was determined in IQTREE and was the Whelan and Goldman (WAG) model, incorporating a discrete gamma distribution (shape parameter = 5) to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). Bootstrap values are indicated next to the branches. A GH5_2 sequence derived from the bacterium *Zobellia galactanivorans* (ZGA, CAZ97841.1) was used as an outgroup. Details on the sequences used here can be found in Table S1). AJA: *A. japonica*; AGL: *A. glabripennis*; AGE: *Apriona germari*; PHI: *P. hilaris*; OAL: *Oncideres albomarginata chamela*; ACH: *Anoplophora chinensis*; MMY: *Mesosa myops*; MAL: *Monochamus alternatus.*

may partly explain the discrepancies we observed. More likely, the presence of the aminopentyl linker at the reducing end of the oligosaccharides produced by automated glycan assembly influences the way the enzymes deal with these substrates relative to the way they deal with oligosaccharides harboring a free reducing end. The potential influence that such a linker present at the reducing end of a synthetic substrate exerts on the catalytic activity of an enzyme could be reduced in the future, thanks to the development of photo-cleavable linkers leading to oligosaccharides with free reducing ends.[15] Alternatively, we also suspect that *AJA*GH5_2-1 could perform some transglycosylation using the xylotetraose. Oddly, we observed xylobiose and xylotriose when *AJA*GH5_2-1 was incubated with xylotetraose, although the enzyme is apparently unable to use xylopentaose as a substrate. Such transglycosylation could also explain the difference in the patterns of breakdown products we observed between the commercially available xylohexaose and the one produced by automated glycan assembly. Transglycosylation is not uncommon for GH5 enzymes[16] and is also influenced by substrate concentration, with transglycosylation being favored at high substrate concentrations compared to hydrolysis.[17] But to be entirely sure that *AJA*GH5_2-1 has the ability to perform transglycosylation in the presence of xylan oligosaccharides, further experiments are required. Until such experiments are performed, this theory remains a hypothesis.

Cerambycid-derived xylanases, such as *AJA*GH5_2-1, *AGL*GH5_2-1 from the Asian long-horned beetle[4b] and others which cluster in clade I (Figure 6), are not unique to insects. Other insect-derived xylanases have been described and merit further functional characterization. Some of these enzymes are part of typical families of xylanases, in particular, two GH11s from the mustard leaf beetle *Phaedon cochleariae*[18] and two GH10s encoded by the genome of the coffee berry borer *Hypothenemus hampei*.[4f,19] Other insect-derived xylanases are part of unexpected GH families. For example, in the bean beetle *Callosobruchus maculatus*, a protein from the subfamily 10 of GH5 (GH5_10)—a typical family of mannanases[11]—has been identified; this protein accepts only xylan as a substrate.[20] In stick and leaf insects (Phasmatodea), a gene family encoding GH9 proteins expanded through several gene duplication events, resulting in one clade of orthologous GH9 proteins having the ability to use both amorphous cellulose and xylan as substrates.[21] We expect that more insect-derived xylanases will be identified in the near future, due to the increasing amount of transcriptome/genome data which are currently generated, in particular for phytophagous/xylophagous species.

As we have shown here, arabinoxylan oligomers generated by automated glycan assembly,[12] together with polysaccharides synthesized by the enzymatic polymerization of arabinoxylan oligosaccharide fluorides,[13] represent powerful tools for the functional characterization of novel xylanases. Such substrates would be helpful to have for analyzing other classes of PCWDEs, such as those acting on substituted polysaccharides like xyloglucan or galactomannan/galactoglucomannan. The synthesis of xyloglucan oligosaccharides by glycan-automated assembly has been developed,[22] and such substrates could be used to functionally characterize insect-derived xyloglucanases such as the GH45 proteins described from several chrysomelid and curculionid species.[3a,23]

Conclusions

We believe that phytophagous insects in general, particularly Phytophaga beetles, represent an abundant source of novel enzymes that can be used to convert plant biomass into fermentable reducing sugars. For too long, the potential of insects as a source of enzymes for biotechnology applications has been overlooked in favor of microorganisms. This neglect is mostly due to the fact that relative to insects, microorganisms are easier to get access to and to cultivate in the labora-

tory. With more than 130000 species, each of whose respective genomes harbor on average more than 10 genes encoding various PCWDEs,[4a,b,d–f,24] Phytophaga beetles represent the largest, under-investigated pool of novel enzymes with potential use in the generation of biofuels.[25]

Experimental Section

Heterologous expression of AJAGH5_2-1 in insect Sf9 cells: We used a construct described in[6d] in which the open reading frame (ORF) of *AJA*GH5_2-1 was cloned in the pIB/V5-His TOPO/TA (Invitrogen), in frame with a V5-(His)6 epitope at the carboxyl-terminus. *Sf*9 cells, cultured in SF-900 II serum-free medium (Gibco, Paisley, UK), were seeded in 6-well plates such that they covered 60 percent of the bottom of each well. These cells were then transfected with the construct described above, using FUGENE HD (Promega, Madison, WI, USA) as the transfection reagent. A mock transfection in which the plasmid DNA was omitted was performed and used as a negative control. After 72 h, the culture medium, which contained the heterologously expressed protein, was harvested and centrifuged (16000 *g*, 5 min, 4 °C) in order to remove floating cells and cell debris. The resulting culture medium was further processed by dialyzing it against 50 mM citrate/phosphate buffer pH 5.0 at 4 °C for 48 h, using Slide-A-Lyzer Dialysis Cassettes with a 10 kDa cutoff, before being desalted with Zeba Desalt Spin Columns 7 kDa cutoff (both Thermo Scientific), according to the manufacturer's guidelines. The resulting crude enzyme extracts were stored at 4 °C until use. Successful expression was verified by Western blot using the anti-V5-HRP antibody.

TLC analysis of hydrolysis end-products of commercially available substrates: Enzyme assays (20 μL) were set up using dialyzed and desalted crude enzyme extracts (14 μL) mixed with a 1% solution of substrate (4 μL) in a 20 mM citrate/phosphate buffer pH 5.0. The substrates used were carboxymethylcellulose (Sigma–Aldrich), regenerated amorphous cellulose (RAC), prepared as described in,[3a] xyloglucan from Tamarind seeds (Megazyme, Bray, Ireland), glucomannan from konjac (Megazyme), galactomannan from carob (Megazyme), beechwood xylan (Sigma–Aldrich), arabinoxylan from rye (Megazyme) and arabinoxylan from wheat (Megazyme). The oligosaccharides used as standards or in enzyme assays were all purchased from Megazyme. Enzyme assays (20 μL) using oligosaccharides as substrates were set up as follows: Dialyzed and desalted crude enzyme extracts (14 μL) were mixed with a given oligosaccharide (0.5 μL; 10 μg μL^{-1}) in a 20 mM citrate/phosphate buffer pH 5.0. Enzyme assays were incubated for 16 h at 40 °C before being applied to TLC plates (silica gel 60, 20×20 cm, Merck). Plates were developed with ethyl acetate/acetic acid/formic acid/water (9:3:1:4) as the mobile phase. Hydrolysis end-products were then revealed by soaking the plates in 0.2 percent (*w*/*v*) orcinol in methanol/sulfuric acid (9:1), then heated briefly until spots appeared on the plates.

Analysis of hydrolysis end-products of synthetic oligo- and polysaccharides: Enzyme assays (20 μL) were set up using dialyzed and desalted crude enzyme extracts (10 μL) of *AJA*GH5_2-1 and a solution of xylan substrates (10 μL) in 100 mM sodium acetate buffer, pH 5. The synthetic oligosaccharides were prepared by automated glycan assembly using an automated glycan synthesizer as previously described[12] and the synthetic polysaccharides were prepared by chemo-enzymatic synthesis.[13] The oligosaccharides were used at a final concentration of 0.5 mM and the artificial xylan polymers were used at a final concentration of 0.75 mg mL^{-1}. All reactions were incubated for 16 h at 40 °C and terminated by incubation for 5 min at 80 °C. The reactions were analyzed on an Agilent 1200 Series HPLC equipped with an Agilent 6130 quadrupole mass spectrometer (MS) and an Agilent 1200 evaporative light-scattering detector (ELSD). The hydrolysis end-products were separated on a Hypercarb column (150×4.6 mm, Thermo Scientific) using a water (including 0.1% formic acid)-acetonitrile (ACN) gradient at a flow rate of 0.7 mL min^{-1} starting at 2.5% ACN for 5 min, ramping up to 5% ACN at 8 min, followed by a slow increase of ACN to 30% at 40 min, a steep ramp to 100 percent ACN at 42 min, a decline back to 2.5 percent ACN from 45 min to 47 min, and equilibration until 55 min at 2.5% ACN. The peaks in the ELSD traces were assigned based on their retention time and the corresponding masses in the MS as determined by Senf et al.[12]

Sequence alignments and phylogenetic analysis: Amino acid sequences of cerambycid-derived GH5_2 proteins were recovered from public databases (Table S1). Amino acid alignments were carried out using MUSCLE version 3.7 implemented in MEGA7[26] and were inspected and corrected manually when needed. Phylogenetic relationships were estimated using maximum likelihood analysis and were conducted using the IQTREE web server.[27] The best model of protein evolution was determined in the IQTREE web server and was determined to be the Whelan and Goldman (WAG) model, incorporating a discrete gamma distribution to model evolutionary rate differences among sites (+G) and a proportion of invariable sites (+I). The robustness of each analysis was tested using 1000 ultrafast bootstrap replicates.

Acknowledgements

We are grateful to Bianca Wurlitzer (MPI-CE, Jena) for technical support. We thank Emily Wheeler, Boston, for editorial assistance. This work was supported by the Max Planck Society. Y.P. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG; PA2808/4-1). F.P. acknowledges support from the DFG (Emmy Noether program PF850/1-1).

Conflict of Interest

The authors declare no conflict of interest.

Keywords: enzymes · GH5_2 · insects · synthetic oligosaccharides · xylanase

[1] a) S. Haddad, D. D. Mckenna, *Syst. Entomol.* **2016**, *41*, 697–716; b) A. E. Marvaldi, C. N. Duckett, K. M. Kjer, J. J. Gillespie, *Zool. Scr.* **2009**, *38*, 63–77; c) D. D. McKenna, A. S. Sequeira, A. E. Marvaldi, B. D. Farrell, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7083–7088.

[2] a) B. D. Farrell, *Science* **1998**, *281*, 555–559; b) D. D. McKenna, A. L. Wild, K. Kanda, C. L. Bellamy, R. G. Beutel, M. S. Caterino, C. W. Farnum, D. C. Hawks, M. A. Ivie, M. L. Jameson, R. A. B. Leschen, A. E. Marvaldi, J. V. Mchugh, A. F. Newton, J. A. Robertson, M. K. Thayer, M. F. Whiting, J. F. Lawrence, A. Slipinski, D. R. Maddison, B. D. Farrell, *Syst. Entomol.* **2015**, *40*, 835–880.

[3] a) A. Busch, G. Kunert, N. Wielsch, Y. Pauchet, *Insect Mol. Biol.* **2018**, *27*, 633–650; b) R. Kirsch, N. Wielsch, H. Vogel, A. Svatos, D. G. Heckel, Y. Pauchet, *BMC Genomics* **2012**, *13*, 587; c) A. M. Scrivener, H. Watanabe, H. Noda, *J. Insect Physiol.* **1997**, *43*, 1039–1052.

[4] a) C. I. Keeling, M. M. Yuen, N. Y. Liao, T. R. Docking, S. K. Chan, G. A. Taylor, D. L. Palmquist, S. D. Jackman, A. Nguyen, M. Li, H. Henderson, J. K. Janes, Y. Zhao, P. Pandoh, R. Moore, F. A. Sperling, D. P. Huber, I. Birol, S. J. Jones, J. Bohlmann, *Genome Biol.* **2013**, *14*, R27; b) D. D.

McKenna, E. D. Scully, Y. Pauchet, K. Hoover, R. Kirsch, S. M. Geib, R. F. Mitchell, R. M. Waterhouse, S. J. Ahn, D. Arsala, J. B. Benoit, H. Blackmon, T. Bledsoe, J. H. Bowsher, A. Busch, B. Calla, H. Chao, A. K. Childers, C. Childers, D. J. Clarke, L. Cohen, J. P. Demuth, H. Dinh, H. Doddapaneni, A. Dolan, J. J. Duan, S. Dugan, M. Friedrich, K. M. Glastad, M. A. Goodisman, S. Haddad, Y. Han, D. S. Hughes, P. Ioannidis, J. S. Johnston, J. W. Jones, L. A. Kuhn, D. R. Lance, C. Y. Lee, S. L. Lee, H. Lin, J. A. Lynch, A. P. Moczek, S. C. Murali, D. M. Muzny, D. R. Nelson, S. R. Palli, K. A. Panfilio, D. Pers, M. F. Poelchau, H. Quan, J. Qu, A. M. Ray, J. P. Rinehart, H. M. Robertson, R. Roehrdanz, A. J. Rosendale, S. Shin, C. Silva, A. S. Torson, I. M. Jentzsch, J. H. Werren, K. C. Worley, G. Yocum, E. M. Zdobnov, R. A. Gibbs, S. Richards, *Genome Biol.* **2016**, *17*, 227; c) Y. Pauchet, C. A. Saski, F. A. Feltus, I. Luyten, H. Quesneville, D. G. Heckel, *Insect Mol. Biol.* **2014**, *23*, 286 – 300; d) Y. Pauchet, P. Wilkinson, R. Chauhan, R. H. Ffrench-Constant, *PLoS One* **2010**, *5*, e15635; e) S. D. Schoville, Y. H. Chen, M. N. Andersson, J. B. Benoit, A. Bhandari, J. H. Bowsher, K. Brevik, K. Cappelle, M. M. Chen, A. K. Childers, C. Childers, O. Christiaens, J. Clements, E. M. Didion, E. N. Elpidina, P. Engsontia, M. Friedrich, I. Garcia-Robles, R. A. Gibbs, C. Goswami, A. Grapputo, K. Gruden, M. Grynberg, B. Henrissat, E. C. Jennings, J. W. Jones, M. Kalsi, S. A. Khan, A. Kumar, F. Li, V. Lombard, X. Ma, A. Martynov, N. J. Miller, R. F. Mitchell, M. Munoz-Torres, A. Muszewska, B. Oppert, S. R. Palli, K. A. Panfilio, Y. Pauchet, L. C. Perkin, M. Petek, M. F. Poelchau, E. Record, J. P. Rinehart, H. M. Robertson, A. J. Rosendale, V. M. Ruiz-Arroyo, G. Smagghe, Z. Szendrei, G. W. C. Thomas, A. S. Torson, I. M. Vargas Jentzsch, M. T. Weirauch, A. D. Yates, G. D. Yocum, J. S. Yoon, S. Richards, *Sci. Rep.* **2018**, *8*, 1931; f) F. E. Vega, S. M. Brown, H. Chen, E. Shen, M. B. Nair, J. A. Ceja-Navarro, E. L. Brodie, F. Infante, P. F. Dowd, A. Pain, *Sci. Rep.* **2015**, *5*, 12525.

[5] R. Kirsch, L. Gramzow, G. Theissen, B. D. Siegfried, R. H. Ffrench-Constant, D. G. Heckel, Y. Pauchet, *Insect Biochem. Mol. Biol.* **2014**, *52*, 33 – 50.

[6] a) N. Calderón-Cortés, H. Watanabe, H. Cano-Camacho, G. Zavala-Paramo, M. Quesada, *Insect Mol. Biol.* **2010**, *19*, 323 – 336; b) C. J. Chang, C. P. Wu, S. C. Lu, A. L. Chao, T. H. Ho, S. M. Yu, Y. C. Chao, *Insect Biochem. Mol. Biol.* **2012**, *42*, 629 – 636; c) J. Liu, K. Song, H. Teng, B. Zhang, W. Li, H. Xue, X. Yang, *Acta Biochim. Biophys. Sin.* **2015**, *47*, 741 – 748; d) Y. Pauchet, R. Kirsch, S. Giraud, H. Vogel, D. G. Heckel, *Insect Biochem. Mol. Biol.* **2014**, *49*, 1 – 13; e) M. Sugimura, H. Watanabe, N. Lo, H. Saito, *Eur. J. Biochem.* **2003**, *270*, 3455 – 3460; f) Y. D. Wei, K. S. Lee, Z. Z. Gui, H. J. Yoon, I. Kim, G. Z. Zhang, X. Guo, H. D. Sohn, B. R. Jin, *Comp. Biochem. Physiol. Part B* **2006**, *145*, 220 – 229.

[7] J. D. Allison, J. H. Borden, S. J. Seybold, *Chemoecology* **2004**, *14*, 123 – 150.

[8] a) E. G. Linsley, *Annu. Rev. Entomol.* **1959**, *4*, 99 – 138; b) M. M. Martin, *Comp. Biochem. Physiol. Part A* **1983**, *75*, 313 – 324.

[9] a) C. Chararas, J. M. Chipoulet, *Comp. Biochem. Physiol. Part B* **1982**, *72*, 559 – 564; b) C. Chararas, J. M. Chipoulet, *Aust. J. Biol. Sci.* **1983**, *36*, 223 – 233; c) C. Chararas, R. Eberhard, J. E. Courtois, F. Petek, *Insect Biochem.* **1983**, *13*, 213 – 218; d) J. M. Chipoulet, C. Chararas, *Comp. Biochem. Physiol. Part B* **1985**, *80*, 241 – 246; e) J. M. Chipoulet, C. Chararas, *Comp. Biochem. Physiol. Part B* **1985**, *82*, 327 – 332; f) M. J. Foglietti, J. E. Courtois, C. Chararas, V. Levaditou, *C. R. Soc. Biol.* **1971**, *165*, 1019 – 1021; g) J. J. Kukor, D. P. Cowan, M. M. Martin, *Physiol. Zool.* **1988**, *61*, 364 – 371; h) J. J. Kukor, M. M. Martin, *J. Chem. Ecol.* **1986**, *12*, 1057 – 1070.

[10] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, *Nucleic Acids Res.* **2014**, *42*, D490-495.

[11] H. Aspeborg, P. M. Coutinho, Y. Wang, H. Brumer 3rd, B. Henrissat, *BMC Evol. Biol.* **2012**, *12*, 186.

[12] D. Senf, C. Ruprecht, G. H. de Kruijff, S. O. Simonetti, F. Schuhmacher, P. H. Seeberger, F. Pfrengle, *Chem. Eur. J.* **2017**, *23*, 3197 – 3205.

[13] D. Senf, C. Ruprecht, S. Kishani, A. Matic, G. Toriz, P. Gatenholm, L. Wagberg, F. Pfrengle, *Angew. Chem. Int. Ed.* **2018**, *57*, 11987 – 11992; *Angew. Chem.* **2018**, *130*, 12163 – 12168.

[14] a) P. Biely, M. Vrsanska, M. Tenkanen, D. Kluepfel, *J. Biotechnol.* **1997**, *57*, 151 – 166; b) A. Pollet, J. A. Delcour, C. M. Courtin, *Crit. Rev. Biotechnol.* **2010**, *30*, 176 – 191; c) M. Vardakou, P. Katapodis, M. Samiotaki, D. Kekos, G. Panayotou, P. Christakopoulos, *Int. J. Biol. Macromol.* **2003**, *33*, 129 – 134.

[15] M. Wilsdorf, D. Schmidt, M. P. Bartetzko, P. Dallabernardina, F. Schuhmacher, P. H. Seeberger, F. Pfrengle, *Chem. Commun.* **2016**, *52*, 10187 – 10189.

[16] a) M. Couturier, A. Roussel, A. Rosengren, P. Leone, H. Stalbrand, J. G. Berrin, *J. Biol. Chem.* **2013**, *288*, 14624 – 14635; b) A. Dilokpimol, H. Nakai, C. H. Gotfredsen, M. J. Baumann, N. Nakai, M. Abou Hachem, B. Svensson, *Biochim. Biophys. Acta* **2011**, *1814*, 1720 – 1729; c) P. Zhou, Y. Liu, Q. Yan, Z. Chen, Z. Qin, Z. Jiang, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *70*, 2970 – 2982.

[17] a) R. Kawai, K. Igarashi, M. Kitaoka, T. Ishii, M. Samejima, *Carbohydr. Res.* **2004**, *339*, 2851 – 2857; b) F. Nanjo, K. Sakai, M. Ishikawa, K. Isobe, T. Usui, *Agr. Biol. Chem. Tokyo* **1989**, *53*, 2189 – 2195; c) D. H. D. Nguyen, J. T. Park, J. H. Shim, P. L. Tran, E. F. Oktavina, T. L. H. Nguyen, S. J. Lee, C. S. Park, D. Li, S. H. Park, D. Stapleton, J. S. Lee, K. H. Park, *J. Bacteriol.* **2014**, *196*, 1941 – 1949.

[18] a) C. Girard, L. Jouanin, *Insect Biochem. Mol. Biol.* **1999**, *29*, 1129 – 1142; b) Y. Pauchet, D. G. Heckel, *Proc. R. Soc. London Ser. B* **2013**, *280*, 20131021.

[19] B. Padilla-Hurtado, C. Florez-Ramos, C. Aguilera-Galvez, J. Medina-Olaya, A. Ramirez-Sanjuan, J. Rubio-Gomez, R. Acuna-Zornosa, *BMC Res. Notes* **2012**, *5*, 23.

[20] A. Busch, G. Kunert, D. G. Heckel, Y. Pauchet, *PLoS One* **2017**, *12*, e0184305.

[21] M. Shelomi, D. G. Heckel, Y. Pauchet, *Insect Biochem. Mol. Biol.* **2016**, *71*, 1 – 11.

[22] P. Dallabernardina, F. Schuhmacher, P. H. Seeberger, F. Pfrengle, *Org. Biomol. Chem.* **2016**, *14*, 309 – 313.

[23] A. Busch, E. G. J. Danchin, Y. Pauchet, *BMC Evol. Biol.* **2019**, *19*, 100.

[24] S. I. Eyun, H. Wang, Y. Pauchet, R. H. Ffrench-Constant, A. K. Benson, A. Valencia-Jimenez, E. N. Moriyama, B. D. Siegfried, *PLoS One* **2014**, *9*, e94052.

[25] C. Luo, Y. Li, Y. Chen, C. Fu, X. Nong, Y. Yang, *Biotechnol. Biofuels* **2019**, *12*, 75.

[26] S. Kumar, G. Stecher, K. Tamura, *Mol. Biol. Evol.* **2016**, *33*, 1870 – 1874.

[27] J. Trifinopoulos, L. T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res.* **2016**, *44*, W232-235.

Manuscript received: November 12, 2019
Accepted manuscript online: December 18, 2019
Version of record online: January 21, 2020

