Functional carbohydrate-active enzymes acquired by horizontal gene transfer from plants in the whitefly *Bemisia tabaci*

Colinet D*[1], Haon M[2,3], Drula E[2,4], Boyer M[1], Grisel S[2,3], Belliardo C[1], Koutsovoulos GD[1], Berrin JG[2,3], Danchin EGJ[1]

[1] Institut Sophia Agrobiotech, INRAE, Université Côte d'Azur, CNRS, Sophia Antipolis, France

[2] Biodiversité et Biotechnologie Fongiques (BBF), INRAE, Aix Marseille Université, Marseille, France

[3] INRAE, Aix Marseille Université, 3PE Platform, Marseille, France

[4] Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix Marseille Université, Marseille, France

* Corresponding author

Correspondence: dominique.colinet@inrae.fr

Summary

Carbohydrate-active enzymes (CAZymes) involved in the degradation of plant cell walls and/or the assimilation of plant carbohydrates for energy uptake are widely distributed in microorganisms. In contrast, they are less frequent in animals, although there are exceptions, including examples of CAZymes acquired by horizontal gene transfer (HGT) from bacteria or fungi in a number of phytophagous arthropods and plant-parasitic nematodes. Although the whitefly *Bemisia tabaci* is a major agricultural pest, knowledge of HGT-acquired CAZymes in this phloem-feeding insect of the Hemiptera order (subfamily Aleyrodinae) is still lacking. We performed a comprehensive and accurate detection of HGT candidates in *B. tabaci* and identified 136 HGT events, 14 of which corresponding to CAZymes. The *B. tabaci* HGT-acquired CAZymes were not only of bacterial or fungal origin, but some were also acquired from plants. Biochemical analysis revealed that members of the glycoside hydrolase families 17 (GH17) and 152 (GH152) acquired from plants are functional beta-glucanases with different substrate specificities, suggesting distinct roles. These two CAZymes are the first characterized GH17

and GH152 glucanases in a metazoan. We identified a lower number of HGT events in the related Aleyrodinae *Trialeurodes vaporariorum*, with only three HGT-acquired CAZymes, including one GH152 glucanase, suggesting a unique HGT event in the ancestor of the Aleyrodinae. A GH152 CAZyme acquired from plants was also identified in two plant cell-feeding insects of the Thysanoptera order, suggesting convergent acquisition and highlighting the importance of plant-acquired CAZymes in the biology of piercing-sucking insects.

Introduction

Horizontal gene transfer (HGT) can be defined as the transfer of genetic material between organisms without conventional vertical transmission from parent to offspring. HGT allows for the transfer of genes between different species, regardless of their evolutionary distance. This phenomenon has been extensively studied in bacteria, for which the supporting mechanisms have been comprehensively documented (Arnold et al. 2022). Although less common than in bacteria, it is becoming increasingly evident that HGT has had a significant impact on the evolution of multicellular eukaryotic genomes, driving functional novelty. In recent years, a number of HGT events have been described between bacteria or viruses and eukaryotes, and even between different eukaryotic organisms (Keeling and Palmer 2008; Crisp et al. 2015; Soucy et al. 2015; Drezen et al. 2017; Husnik and McCutcheon 2018; Chen et al. 2021). Among eukaryotes, HGT has played an important role in the ability of many animals to feed on plants, including phytophagous arthropods (Kirsch et al. 2014; Nakabachi 2015; Wybouw et al. 2016; Husnik and McCutcheon 2018) and plant-parasitic nematodes (Haegeman et al. 2011; Danchin and Rosso 2012; Husnik and McCutcheon 2018). In particular, a number of genes encoding carbohydrate-active enzymes (CAZymes) involved in plant cell wall degradation have been described as horizontally transferred in arthropods and nematodes.

The plant cell wall is a dynamic structure that serves multiple functions, including protection from biotic stresses. It is composed of a network of high molecular weight polysaccharides, including β-1,4 glucans (cellulose), β-1,4 xylans, β-1,4 mannans or mixed β-1,3/β-1,4 glucans (hemicelluloses), and the heteropolysaccharide pectin (Perrot et al. 2022). The first plant cell wall degrading enzymes (PCWDEs) identified in an animal were ß-1,4-endoglucanases (or cellulases), which belong to the

57 glycoside hydrolase 5 (GH5) family of CAZymes (Drula et al. 2022) and were presumably horizontally
58 acquired from bacteria by plant-parasitic nematodes (Smant et al. 1998). Since then, numerous
59 examples of CAZymes acquired by HGT from bacterial or fungal donors and active on polysaccharides
60 constitutive of the plant cell wall have been described in plant-parasitic nematodes (Danchin et al. 2010;
61 Haegeman et al. 2011; Danchin and Rosso 2012; Palomares-Rius et al. 2014; Vicente et al. 2019), but
62 also in phytophagous arthropods (Pauchet et al. 2010; Pauchet and Heckel 2013; Kirsch et al. 2014;
63 Pauchet et al. 2014; Evangelista et al. 2015; Antony et al. 2017; Faddeeva-Vakhrusheva et al. 2017;
64 Busch et al. 2019; Shin et al. 2021; Le et al. 2022; Shin and Pauchet 2023). HGT-acquired CAZymes
65 identified in plant-parasitic nematodes and phytophagous arthropods also include enzymes directly
66 involved in plant carbohydrate assimilation for energy metabolism (Acuña et al. 2012; Sun et al. 2013;
67 Danchin et al. 2016; Dai et al. 2021).

68 The whitefly *Bemisia tabaci* (Hemiptera: Aleyrodinae) is a major agricultural pest. It feeds on
69 phloem sap, causing damage directly through feeding with its piercing-sucking mouthparts and
70 indirectly through the transmission of numerous plant pathogenic viruses (Navas-Castillo et al. 2011).
71 *B. tabaci* is a highly polyphagous species complex of more than 30 cryptic species, among which the
72 genome of the Middle East-Asia Minor 1 (MEAM1), Mediterranean (MED), and SSA-East and Central
73 Africa (SSA-ECA) pests have been recently released (Chen et al. 2016; Xie et al. 2017; Chen et al.
74 2019). HGT appears to be widespread in *B. tabaci,* with numerous genes reported to have been
75 transferred not only from bacteria or fungi (Chen et al. 2016; Li et al. 2022), but also from plants (Gilbert
76 and Maumus 2022; Li et al. 2022), making *B. tabaci* the first metazoan species documented to have
77 acquired genes of plant origin (Lapadula et al. 2020). Among these horizontally acquired genes,
78 functional evidence supports the involvement of genes of bacterial origin in B vitamin synthesis (Ren
79 et al. 2020) and nitrogen metabolism (Yang et al. 2024), genes of fungal origin in the ferredoxin-
80 mediated suppression of plant defenses (Wang et al. 2023), and genes of plant origin in processes such
81 as detoxification of plant toxic compounds (Xia et al. 2021) or reproduction (Gong et al. 2023). Careful
82 mining of previous analyses (Chen et al. 2016; Gilbert and Maumus 2022; Li et al. 2022) suggests the
83 possible acquisition of several CAZymes of different origin in *B. tabaci*. However, none have been
84 biochemically characterized and their possible role in the biology of the insect remains unknown.

85 Furthermore, comprehensive and accurate detection of HGT-acquired CAZymes in *B. tabaci* is still
86 lacking, leaving a knowledge gap. Phloem feeders, such as aphids or whiteflies, rely on CAZymes to
87 facilitate stylet penetration through plant cell walls into a phloem sieve element, but also to counteract
88 the sieve element occlusion defense mechanism consisting of callose (β-1,3-glucan) deposition (Silva-
89 Sanzana et al. 2020; Walker 2022). CAZymes are also needed to transform sugars taken up from the
90 plant, mainly sucrose, into fructose for energy metabolism. Finally, some phloem feeders depend on
91 CAZymes to overcome the high osmolarity of the phloem through osmoregulation (Douglas 2006).
92 In this work, we used comprehensive phylogenetic analyses to identify HGT events in *B. tabaci*
93 MEAM1 that are specific to the Aleyrodinae, and we performed further characterization with a focus
94 on CAZymes. Biochemical analysis revealed that two of the plant-acquired CAZymes, belonging to the
95 GH17 and GH152 families, are functional beta-glucanases with different substrate specificities,
96 suggesting distinct functional roles. We then performed a comparative analysis with the related
97 greenhouse whitefly *Trialeurodes vaporariorum* (Hemiptera: Aleyrodinae), which differs from *B.*
98 *tabaci* in various aspects such as host plant range, virus transmission, or insecticide resistance. Finally,
99 we expanded our analysis beyond the Aleyrodinae to detect HGT in two other piercing-sucking insects,
100 *Frankliniella occidentalis* and *Thrips palmi* (Thysanoptera: Thripinae), which feed on the contents of
101 plant cells. Overall, our findings suggest that the acquisition of CAZymes by HGT has had a significant
102 impact on the biology of piercing-sucking insects, particularly in the case of *B. tabaci*.
103

104 **Material and Methods**

105 *Data used and quality control*

106 Genome and predicted proteome data from the *Bemisia tabaci* MEAM1 (v1.2) (Chen et al.
107 2016), MED (v1.0) (Xie et al. 2017), and SSA-ECA (v1.0) (Chen et al. 2019) cryptic species and from
108 *Trialeurodes vaporariorum* (v1.0) (Xie et al. 2020) were obtained from the Whitefly Genome Database
109 (http://www.whiteflygenomics.org).

110 Genome and predicted proteome data from *Frankliniella occidentalis* (assembly
111 GCF_000697945.2_Focc_2.1) and *Thrips palmi* (assembly GCF_012932325.1_TpBJ-2018v1) were
112 obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/).
113 Proteome completeness was compared between *B. tabaci* MEAM1, MED and SSA-ECA
114 cryptic species, *T. vaporariorum*, *F. occidentalis* and *T. palmi* using BUSCO (v5) according to the
115 Arthropoda Odb10 dataset (Manni et al. 2021).

116 *Orthogroup inference*

117 Orthogroups (i) between the *B. tabaci* MEAM1, MED and SSA-ECA cryptic species and *T.*
118 *vaporariorum* and (ii) between *F. occidentalis* and *T. palmi* were defined using Orthofinder (v2) with
119 default parameters (Emms and Kelly 2015; Emms and Kelly 2019).

120 *Functional annotation*

121 All the predicted proteins from the *B. tabaci* MEAM1, *T. vaporariorum*, *F. occidentalis* and *T.*
122 *palmi* genomes were analyzed using InterProScan (v5) to identify conserved protein domains (Jones et
123 al. 2014). The -iprlookup and -goterms options were used to assign Gene Ontology (GO) terms from
124 the identified InterPro domains.

125 *Detection of HGT candidates*

126 A homology search against the NCBI non-redundant (NR) protein database was performed for
127 the *B. tabaci* MEAM1, *T. vaporariorum*, *F. occidentalis* and *T. palmi* proteomes using DIAMOND (v2)
128 (Buchfink et al. 2021). The DIAMOND search was performed in the more sensitive mode with an e-
129 value threshold of $1.0e^{-3}$ and a maximum number of hits of 500.
130 The DIAMOND homology search results were submitted to AvP (Koutsovoulos et al. 2022) to
131 calculate the Aggregate Hit Score (AHS) (Koutsovoulos et al. 2022) for each query sequence based on
132 the normalized sum of the scores of the best metazoan (NCBI:txid33208) and non-metazoan hits. This
133 AHS metric was shown to be less sensitive to contamination and taxonomic assignment errors in

134 databases than the classical Alien Index (AI) (Koutsovoulos et al. 2022), which only considers the single
135 best metazoan and single best non-metazoan hits. Proteins with an AHS greater than 0 have a higher
136 similarity to non-metazoan than to metazoan hits in the NR database. Self-hits to Aleyrodinae
137 (NCBI:txid33379) were ignored for *B. tabaci* MEAM1 and *T. vaporariorum.* Self-hits to Thysanoptera
138 (NCBI:txid30262) have been ignored for *F. occidentalis* and *T. palmi.*

139 We again used AvP for automatic phylogenetic detection of HGT candidates among proteins
140 with AHS above 0. The first step of AvP was to cluster the query sequences based on the percentage of
141 shared hits in the DIAMOND homology search result (default 70% overlap). This clustering method
142 was supplemented with orthogroups inferred by Orthofinder (see above) and shared Pfam domains as
143 determined by InterProScan (see above). When analyzing the AVP results, we retained the most
144 comprehensive clustering method, for which only one HGT event was predicted from the obtained
145 phylogeny. All protein sequences with significant hits in the DIAMOND homology search result were
146 retrieved from the NR database by AvP and aligned using MAFFT (v7) with the --auto option (Katoh
147 and Standley 2013) for each group. The second step of AvP was to infer the phylogeny for each group
148 and to detect HGT candidates according to the species found in the sister branch of the query sequence
149 and the ancestral sister branch. FastTree (v2) with the default parameters (Price et al. 2010) was
150 preferred to IQ-TREE (v2) (Minh et al. 2020) for phylogeny inference in this initial step to improve the
151 speed of the analysis. In a third step, AvP classified the HGT candidates according to their putative
152 origin. The fourth step of AvP was to infer a constrained topology in which the query sequence(s) and
153 the other metazoan sequence(s), if present, form a single monophyletic group and to determine whether
154 the topology supporting HGT is significantly more likely than the constrained alternative one using an
155 approximately unbiased (AU) test (Shimodaira 2002). Finally, AvP analyzed the genomic environment
156 of each HGT candidate gene and calculated a local score ranging from -1 to +1 depending on whether
157 the gene was surrounded by other HGT candidate genes (indicating a possible contamination,
158 duplications after an HGT event, or multiple HGTs) or not. The local score was not reported if the
159 number of genes in the scaffold (including the HGT candidate) was less than 5.

160 *Validation of HGT candidates*

161 A manual analysis of the AvP results was performed, and HGT candidates were not further
162 considered if at least one of the following criteria was met:

163 (i) The total number of donor sequences was less than 3 in the sister branch plus the
164 ancestral sister branch (if present).

165 (ii) The HGT-supporting topology was not significantly more likely than the constrained
166 alternative topology in which the query was forced to group with all metazoan
167 sequences from NR. However, HGT candidates were retained if these metazoan
168 sequences were either (i) taxonomically misannotated, likely due to contamination by a
169 donor species or (ii) suspected to have originated from an HGT event themselves based
170 on BLASTP results performed at NCBI against NR (https://www.ncbi.nlm.nih.gov/).

171 (iii) At least one donor sequence in the DIAMOND homology search results shared more
172 than 70% identity with the HGT candidate sequence and either (i) no homologous
173 sequence was found in the closely related cryptic species for *B. tabaci* MEAM1,
174 suggesting contamination by a donor species, or (ii) the donor sequences were suspected
175 to be misannotated, probably in the case of Viridiplantae due to contamination of the
176 donor species by the query species (or a related insect species), based on BLASTP
177 results performed at NCBI against NR (https://www.ncbi.nlm.nih.gov/).

178 (iv) The identity between the donor sequences and the HGT candidate sequence in the
179 DIAMOND homology search results was less than 30%. Searching for homologous
180 proteins and building a phylogenetic tree below this value is problematic due to the low
181 quality of the alignments.

182 (v) The average alignment length in the DIAMOND homology search results was less than
183 100 amino acids with an average query coverage of less than 50%.

184 (vi) The local score calculated by AvP from the genomic environment of the HGT candidate
185 was less than 0, and there was no indication of duplication after an HGT event or of
186 multiple HGTs, indicating a possible contamination.

187 *Phylogeny reconstruction for validated HGT candidates*

188 For each HGT candidate detected by AvP and satisfying all the above-mentioned criteria, we
189 reconstructed a maximum likelihood phylogeny with IQ-TREE (v2) (Minh et al. 2020) to improve
190 accuracy. Before reconstructing the phylogeny, the AvP-defined groups were refined by the following
191 three preliminary steps:

192 (i) We searched for homologous sequences of each HGT candidate validated for *B. tabaci*
193 MEAM1 in the cryptic species MED and SSA and included them in the groups formed
194 by AvP. The search for homologous sequence(s) in the cryptic species MED and SSA
195 was first performed in the orthogroups determined by Orthofinder (see above). If no
196 homologous sequence was found, we performed a protein-to-genome comparison using
197 exonerate (v2) with a score threshold of 500 (Slater and Birney 2005). The *B. tabaci*
198 MEAM1 HGT candidate protein sequence was used as the query and the genome of
199 each cryptic species as the database. When a significant hit was found, the protein
200 sequence was inferred from the cryptic species genomic sequence. The homologous
201 sequence(s) found for the cryptic MED and SSA species were added to the *B. tabaci*
202 MEAM1 groups previously defined by AvP.

203 (ii) We combined the groups defined by AvP for *B. tabaci* MEAM1 and *T. vaporariorum* on
204 the one hand and *F. occidentalis* and *T. palmi* on the other hand, where the sequences
205 of the two related species were in the same orthogroup defined by Orthofinder. If no
206 homologous sequence was found for a validated HGT candidate in the related species,
207 a protein-to-genome comparison was performed using exonerate (v2) as described
208 above.

209 (iii) Metazoan sequences suspected to be misannotated, probably due to contamination, based
210 on BLASTP results performed at NCBI against NR (https://www.ncbi.nlm.nih.gov/)
211 were removed from the groups.

212 For each group, a CD-HIT analysis was then performed with an identity threshold of 100% to
213 remove redundancy (Li and Godzik 2006). Sequences were aligned using MAFFT (v7) with the --auto

214 option (Katoh and Standley 2013). Poorly aligned regions were removed using trimal (v1.4) with the -
215 automated1 option (Capella-Gutiérrez et al. 2009). Phylogenies were inferred using IQ-TREE (v2)
216 (Minh et al. 2020) with automated model selection (Kalyaanamoorthy et al. 2017). Support values were
217 based on a Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) (Guindon et al. 2010)
218 combined with an ultrafast bootstrap (UFboot) approximation with 1,000 replicates (Hoang et al. 2018).
219 Only support values greater than or equal to 80% and 95% for SH-aLRT and UFboot, respectively,
220 were considered. Phylogenies were visualized using iTol (Letunic and Bork 2021).

221 When *B. tabaci* MEAM1 and *T. vaporariorum* sequences on the one hand and *F. occidentalis*
222 and *T. palmi* sequences on the other hand did not form monophyletic groups, we forced a constrained
223 topology supporting monophyly of metazoan sequences and determined whether the unconstrained
224 topology was significantly more likely than the constrained alternative one using an approximately
225 unbiased (AU) test (Shimodaira 2002) with IQ-TREE (v2).

226 *Identification of overrepresented GO terms among validated HGT candidates*

227 Identification of overrepresented Gene Ontology (GO) terms among validated HGT candidates
228 was performed using a hypergeometric test as implemented in func (Prüfer et al. 2007) within the R
229 package GOfuncR with a family-wise error rate threshold of 0.05. The -refine option was used to
230 eliminate redundancy between GO terms and to keep only representative terms.

231 *Detection of encoded Carbohydrate-Active enZymes (CAZymes)*

232 All the predicted proteins from the, *T. vaporariorum*, *F. occidentalis* and *T. palmi* genomes
233 were compared with entries in the CAZy database (Drula et al., 2022). A homemade pipeline combining
234 the BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and HMMER3 (http://hmmer.janelia.org/) tools
235 was used to compare protein models with the sequences of the CAZy modules. Proteins with E-values
236 less than 0.1 were further screened by a combination of BlastP searches against libraries generated from
237 the sequences of the catalytic and non-catalytic modules. HMMER3 was used to search against a
238 collection of custom Hidden Markov Model (HMM) profiles constructed for each CAZy family. Expert
239 curators then performed manual inspection to resolve borderline cases.

240 *Structural analysis*

241 Structure prediction was performed using the ColabFold v1.5.2 implementation of AlphaFold2
242 (Mirdita et al. 2022) at
243 https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
244 with the following parameters: num_relax=1 and template_mode=pdb100. Secondary structure
245 assignment was performed with the DSSP program (https://www3.cmbi.umcn.nl/xssp/) (Kabsch and
246 Sander 1983). Visualization of predicted protein structures, alignment of the alpha carbon atoms with
247 crystallized structures, and calculation of the root-mean-square deviation (RMSD) value were
248 performed using PyMOL (http://sourceforge.net/projects/pymol/).

249 *Recombinant production of CAZymes*

250 The recombinant proteins were produced using the in-house 3PE Platform (*Pichia Pastoris*
251 Protein Express; www.platform3pe.com/) as described in Haon et al. (2015). The sequences of the genes
252 coding for the GH17 Bta06115 and the GH152 Bta13961 from *B. tabaci* were synthesized after codon
253 optimization for expression in *Pichia pastoris* and inserted into a modified pPICZαA vector using *Xho*I
254 and *Xba*I restriction sites in frame with the α secretion factor at the N terminus (i.e., without native
255 signal peptide) and with a $(His)_6$-tag at the C terminus (without *c*-myc epitope) (Genewiz, Leipzig,
256 Germany). Transformation of competent *P. pastoris* X33 cells (Invitrogen, Carlsbad, California, USA)
257 was performed by electroporation with PmeI-linearized pPICZαA recombinant plasmids as described
258 in Haon et al. (2015). Zeocin-resistant transformants were then screened for protein production.

259 *Heterologous protein production in flasks*

260 The best-producing transformants (GH17 Bta06115 and GH152 Bta13961) were grown in 2 L
261 BMGY medium (10 $g.L^{-1}$ glycerol, 10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ peptone, 3.4 $g.L^{-1}$ YNB, 10 $g.L^{-1}$
262 ammonium sulfate, 100 mM phosphate buffer pH 6 and 0.2 $g.L^{-1}$ of biotin) at 30°C and 200 rpm to an
263 optical density at 600 nm of 2-6. Expression was induced by transferring cells into 400 mL of BMMY
264 media at 20 °C in an orbital shaker (200 rpm) for another 3 days. Each day, the medium was

265 supplemented with 3% (v/v) methanol. The cells were harvested by centrifugation, and just before
266 purification, the pH was adjusted to 7.8 and was filtered on 0.45-µm membrane (Millipore, Burlington,
267 Massachusetts, USA).

268 *Purification by affinity chromatography*

269 Filtered culture supernatant was loaded onto a 20 ml HisPrep FF 16/10 column (Cytiva, Vélizy-
270 Villacoublay, France) equilibrated with buffer A (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole
271 10 mM) that was connected to an Äkta pure (Cytiva). $(His)_6$-tagged recombinant proteins were eluted
272 with buffer B (Tris-HCl 50 mM pH 7.8, NaCl 150 mM, imidazole 250 mM). Fractions containing
273 recombinant enzymes were pooled, concentrated, and dialyzed against Tris-HCl 50 mM pH 7.8, NaCl
274 150 mM.

275 Concentration of purified protein was determined by absorption at 280 nm using a Nanodrop
276 ND-2000 spectrophotometer (Thermo Fisher Scientific) with calculated molecular mass and molar
277 extinction coefficients derived from the sequences. Proteins were loaded onto a 10% Tris-glycine
278 precast SDS-PAGE gel (BioRad, Gemenos, France) which was stained with Coomassie Blue. The
279 molecular mass under denaturing conditions was determined with reference standard proteins (Page
280 Ruler Prestained Protein Ladder, Thermo Fisher Scientific, Waltham, MA, USA).

281 *Functional enzymatic assays*

282 Enzyme assays (final reaction volume 200 µL) were performed at 30°C in 2 mL Eppendorf
283 tubes, incubated in a thermomixer (Eppendorf, Hamburg, Germany) at 1000 rpm for 24 h in sodium
284 phosphate buffer (50 mM, pH 7.0). Polysaccharides (Barley β-glucan, Lichenan) (MEGAZYME, Bray,
285 Ireland) were used at 10 $mg.mL^{-1}$, and oligosaccharides (laminarihexaose, cellohexaose)
286 (MEGAZYME) at a concentration of 0.5 mM. Reactions were initiated by the addition of Bta06115 and
287 Bta13961 at a final concentration of 8 µM. Reactions were stopped by heating at 100°C for 5 min,
288 centrifuged (12,000*xg*, 2 min, 4°C), and diluted 10-fold in milliQ H_2O before injection onto the HPAEC
289 column.

290 *HPAEC-PAD analyses*

291 The detection method is performed using a high-performance anion-exchange chromatography
292 (HPAEC) coupled with pulsed amperometric detection (PAD) (DIONEX ICS6000 system, Thermo
293 Fisher Scientific). The system is equipped with a CarboPac-PA1 guard column (2 x 50 mm) and a
294 CarboPac-PA1 column (2 x 250 mm) kept at 30°C. Elution was carried out at a flow rate of 0.25
295 $mL.min^{-1}$ and 25 µL of samples was injected. The eluents used were 100 mM NaOH (eluent A) and
296 NaOAc (1M) in 100 mM NaOH (eluent B). The initial conditions were set to 100% eluent A, and the
297 following gradient was applied: 0 to 10 min, 0 to 10% B; 10 to 35 min, 10 to 35% B (linear gradient);
298 35 to 40 min, 30 to 100% B (curve 6); 40 to 41 min, 100 to 0% B; 41 to 50 min, 100% A. Integration
299 was performed using the Chromeleon 7.2.10 data software.

300

301 **Results and Discussion**

302 *The genome of B. tabaci contains numerous genes of non-metazoan origin*

303 Our first objective was to perform a comprehensive and accurate phylogenetic detection of HGT
304 candidates in *B. tabaci,* a member of the Aleyrodinae subfamily in the Hemiptera order, and to compare
305 the results obtained with the literature. The proteome completeness estimated by BUSCO was compared
306 between the three *B. tabaci* cryptic species, MEAM1, MED and SSA-ECA, for which genome and
307 predicted proteome data are available (Supplementary Table 1). With 95.6% of arthropoda BUSCO
308 proteins found to be complete, *B. tabaci* MEAM1 had the most complete proteome, compared to less
309 than 90% for the other two *B. tabaci* cryptic species (Supplementary Table 1). Therefore, the search for
310 candidate HGTs was performed for *B. tabaci* MEAM1 as a reference.

311 Of the 15,662 *B. tabaci* MEAM1 predicted proteins, 686 returned an AHS greater than 0,
312 meaning that they are more similar to non-metazoan proteins than to metazoan proteins. From these
313 predicted proteins, AvP identified 357 as possible HGT candidates specific to Aleyrodinae. The vast
314 majority of these (255 out of 357) were validated according to the criteria described in the Material and
315 Methods section and were from potential bacterial, fungal, or plant (Viridiplantae) donors (Table 1 and

316 Supplementary Table 2). These 255 validated HGT candidates correspond to at least 136 distinct HGT
317 events (Supplementary Table 2). For each of these events, homologs were found in at least one of the
318 other two *B. tabaci* cryptic species, MED and SSA-ECA, making the possibility of contamination
319 unlikely (Supplementary Table 2). Further supporting the lack of contamination, the local score
320 calculated by AvP was greater than 0 for each of the 255 validated HGT candidates. This result (136
321 events for 255 genes) also suggests that most of the candidate HGTs were inserted into multiple
322 independent genomic regions and did not undergo massive cis-duplication (Supplementary Table 2).
323 We compared the HGT candidates of bacterial, fungal and plant origin we found in *B. tabaci*
324 MEAM1 with data from the literature. Most of the HGT candidates described in the literature for *B.*
325 *tabaci* MEAM1 were confirmed by our study suggesting that our approach is comprehensive despite
326 the stringent filters applied (Figure 1). Of the 93 HGT candidates of bacterial origin described in the
327 literature (Chen et al. 2016; Li et al. 2022), 11 were not found in our work (Figure 1 and Supplementary
328 Table 3). Of these, one was no longer present in version 1.2 of the *B. tabaci* MEAM1 proteome, one
329 was rejected because its sequence identity with the donor sequences was below 30%, and one was
330 considered non-HGT as metazoan sequences were present in the sister and ancestral sister branches
331 (Supplementary Table 3). The remaining eight were found to have an AHS score equal to or less than
332 0, indicating a higher similarity to metazoans than to non-metazoans in the NR database and thus
333 unlikely acquisition via HGT (Supplementary Table 3). On the other hand, we identified 31 new
334 phylogenetically supported HGT candidates with bacteria as potential donors not yet described in the
335 literature (Figure 1 and Supplementary Table 2). 17 of these new HGT candidates belong to the AAA-
336 ATPase-like domain-containing protein family, a large family of ATPases associated with various
337 cellular activities. All but one of them group with previously identified HGT candidates in *B. tabaci*
338 MEAM1, suggesting that they originate from the same HGT events but were overlooked in previous
339 analyses. The remaining one would correspond to a new, independent HGT event of a member of the
340 AAA-ATPase-like domain-containing protein family in *B. tabaci* (Supplementary Table 2). Horizontal
341 gene transfer from bacteria has been described for members of this protein family in two other phloem-
342 feeding insects of the order Hemiptera, but belonging to the superfamilies Psylloidea and Coccoidea,

343 separated from Aleyrodinae by more than 250 million years (Myr) (Misof et al. 2014), with a possible
344 role in mediating the interaction with endosymbionts (Sloan et al. 2014).

345 Of the 81 HGT candidates of fungal origin described in the literature (Chen et al. 2016), only
346 two were not retrieved in our work (Figure 1). For one of these two HGT candidates, no similarities
347 were found in the NR database, preventing the search for a possible HGT origin (Supplementary Table
348 3). For the second candidate, classified as HGT_complex in our study, the HGT donor could not be
349 unambiguously determined because the sister branch consists of bacterial sequences and the ancestral
350 sister branch consists of fungal sequences (Supplementary Table 3). On the other hand, we identified
351 nine new phylogenetically confirmed HGT candidates with fungi as potential donors (Figure 1 and
352 Supplementary Table 2). Of these nine new HGT candidates, five cluster with HGT candidates already
353 described in the literature, suggesting that they originated from the same HGT events as the latter
354 (Supplementary Table 2). The other four would correspond to previously undescribed HGT acquisitions
355 of two different genes encoding replication factor A protein 1 and a protein of unknown function,
356 respectively (Supplementary Table 2).

357 Finally, of the 53 HGT candidates of plant origin described in the literature (Lapadula et al.
358 2020; Xia et al. 2021; Gilbert and Maumus 2022; Li et al. 2022), only one was not found in our work
359 (Figure 1). No significant similarities were found in the NR database for this HGT candidate, which
360 prevented the search for a possible HGT origin (Supplementary Table 3). In the work of Gilbert and
361 Maumus (2022), this HGT candidate was identified as a member of a large family of omega-6 fatty acid
362 desaturases, all derived from a single HGT event. The other 14 members of this family were
363 successfully found in our work (Supplementary Table 2). On the other hand, we identified one new
364 HGT candidate acquired from plants (Figure 1 and Supplementary Table 2). This new HGT candidate
365 belongs to the HXXXD family of acyl-transferases. Other members of this family have already been
366 described as HGT candidates for *B. tabaci* MEAM1 (Gilbert and Maumus 2022; Li et al. 2022). This new
367 case would correspond to an independent HGT event in the same family (Supplementary Table 2).

368 Overall, our approach to identify HGT candidates using AvP appeared to be not only reliable, but
369 also more exhaustive than previous approaches used for bacterial and fungal donors (Chen et al. 2016; Li et

370 al. 2022) (Figure 1). For HGT candidates with plants as donors, our approach is comparable to that of Gilbert
371 and Maumus (2022), and both of them are more comprehensive than that of Li et al. (2022) (Figure 1).

372 *B. tabaci has acquired CAZymes from bacteria, fungi, and plant donors via multiple HGT*
373 *events*

374 We identified significantly overrepresented GO terms among validated HGT candidates from
375 potential bacterial, fungal, or plant donors in *B. tabaci* based on their InterPro domain annotation
376 (Supplementary Table 4). This revealed a significant enrichment of GO terms related to carbohydrate
377 metabolism and plant cell wall degradation, such as hydrolase activity, acting on glycosyl bonds
378 (GO:0016798), pectinesterase activity (GO:0030599), or cell wall modification (GO:0042545). This
379 suggests horizontal acquisition of a substantial proportion of Carbohydrate-Active enZymes
380 (CAZymes) in *B. tabaci.* We therefore investigated the relationships between HGT candidates and
381 CAZymes, as well as possible donors, in more detail. A total of 433 *B. tabaci* MEAM1 proteins were
382 predicted to be CAZymes (Supplementary Table 5). These include two members of the Glycoside
383 Hydrolase (GH) family GH13, subfamily 17 (GH13_17), which have been shown to catalyze the
384 detoxification of glucosinolates produced by plants against herbivores (Malka et al. 2020), but were not
385 detected as having been acquired by HGT.

386 We identified 25 CAZymes likely acquired by HGT via 14 independent events (Table 2 and
387 Supplementary Figure 1). Three CAZymes were predicted to originate from bacteria (corresponding to
388 two HGT events), twelve from fungi (corresponding to four HGT events), and nine from plants
389 (corresponding to seven HGT events). For the last CAZyme HGT candidate, it was not possible to
390 decipher whether the origin was bacterial or fungal (Table 2 and Supplementary Figure 1).

391 Two of the CAZymes of bacterial origin belong to the GH32 family and would correspond to a
392 single unique HGT event according to the AvP results (Table 2 and Supplementary Figure 1).
393 Enzymatic activities in the GH32 family include invertase (or ß-fructofuranosidase) activity, which
394 catalyzes the hydrolysis of sucrose into fructose and glucose. Independent horizontal acquisition events
395 of ß-fructofuranosidase genes from bacterial donors have been evidenced for a number of metazoans
396 with a role in assimilation of plant carbohydrates (Nakabachi 2015; Danchin et al. 2016; Wybouw et al.

397 2016; Dai et al. 2021). Acquisition of these GH32 CAZymes from bacteria would have allowed *B.*
398 *tabaci* to metabolize host-derived sucrose, which is abundant in the phloem on which it feeds. The third
399 protein of bacterial origin classified as a CAZyme contains a carbohydrate-binding module of the
400 CBM50 family, but no associated catalytic module (Table 2 and Supplementary Figure 1).

401 Among the CAZymes of fungal origin, three belong to the GH30, GH49 and GH71 families,
402 respectively (Table 2 and Supplementary Figure 1). Interestingly, GH30 enzymes, presumably acquired
403 via HGT, have been found in numerous plant-parasitic nematodes (Danchin et al. 2010), with xylanase
404 activity confirmed in root-knot nematodes (Mitreva-Dautova et al. 2006). However, although xylanase
405 enzymes have been found in other insects, none of those characterized belong to the GH30 family
406 (Padilla-Hurtado et al. 2012; Pauchet and Heckel 2013; Pauchet et al. 2014; Vega et al. 2015). This
407 suggests a convergent acquisition of xylanases from different sources. The GH49 enzyme may be active
408 on plant arabinan, although arabinanase activity has never been confirmed so far in any eukaryote in
409 this GH family. Finally, the only known activity to date in the GH71 family is α-1,3-glucanase, which
410 has been characterized for this family only in fungi. A multigenic family of nine proteins containing a
411 carbohydrate binding module (CBM32) in tandem with a galactose oxidase module (AA5_2) was also
412 found to be horizontally acquired from fungi (Table 2 and Supplementary Figure 1). The biological role
413 of fungal AA5_2 galactose oxidases is not known but other members of this subfamily displaying
414 alcohol oxidase activity play a role in the mechanism of plant penetration in fungal phytopathogens
415 (Bissaro et al. 2022). To the best of our knowledge, there have been no reports of horizontal acquisition
416 of genes containing this combination of modules (AA5_2-CBM32) in metazoans.

417 The HGT candidate whose bacterial or fungal origin could not be determined from the
418 phylogeny belongs to the PL1_4 family (Table 2 and Supplementary Figure 1). PL1_4 enzymes are
419 pectin lyases involved in the degradation of pectin and are usually found in bacteria or fungi
420 (http://www.cazy.org/; Drula et al. 2022). Sequences from a few other insects, including the Thripinae
421 *F. occidentalis* and *T. palmi*, are found in the phylogeny, in different branches for some (Supplementary
422 Figure 1). This will be further investigated in the last part of the Results and Discussion section.

423 As far as we know, no CAZyme acquired by HGT from a plant donor has yet been fully
424 described and characterized in a metazoan. Here, we identified a total of nine *B. tabaci* MEAM1

425 CAZymes, representing seven HGT events of plant origin (Table 2 and Supplementary Figure 1). Of
426 these, three belong to the GlycosylTransferase (GT) families GT10, GT17 and GT61. Of the remaining
427 six, three belong to the GH17 and GH152 families, two belong to the Carbohydrate Esterase 8 (CE8)
428 family, and one is distantly related to plant expansins. The possible functions of some of these
429 CAZymes are described in more detail below.

430 *A GH17 CAZyme acquired by HGT from plants in B. tabaci is a functional β-1,3-glucanase*

431 The *B. tabaci* MEAM1 proteins Bta06115 and Bta06118 share similarities with plant ß-
432 glucanases belonging to the GH17 family (Table 2 and Supplementary Table 2). The GH17 family is
433 well described, with over 50 enzymes biochemically characterized (http://www.cazy.org/; Drula et al.
434 2022). The GH17 family includes two major groups of ß-glucanases, endo-β-1,3-glucanases, which
435 hydrolyze internal β-1,3 glycosidic linkages in β-1,3-glucans, and endo-β-1,3-1,4-glucanases, which
436 hydrolyze internal β-1,4 glycosidic linkages in mixed β-1,3-1,4-glucans (http://www.cazy.org/; Drula
437 et al. 2022). Members of the GH17 family are found in bacteria, fungi and plants, but to our knowledge,
438 they have not been documented in animals so far. Accordingly, homologs of the *B. tabaci* MEAM1
439 Bta06115 and Bta06118 proteins were identified only in *B. tabaci* MED and SSA-ECA, and no other
440 organism besides plants (Figure 2A and Supplementary Table 2). The close proximity of both Bta06115
441 and Bta06118 genes on the same scaffold, together with the bootstrap-supported grouping of all *B.*
442 *tabaci* GH17 sequences (Figure 2A), suggests that they originated from a single HGT event from a plant
443 donor followed by a tandem duplication. The bootstrap-supported grouping of *B. tabaci* MED and SSA-
444 ECA proteins with each of *B. tabaci* MEAM1 Bta06115 and Bta06118 proteins, respectively (Figure
445 2A), further suggests that the duplication occurred before the divergence of the three cryptic species,
446 which is estimated to have occurred between approximately 5 Myr ago (Chen et al. 2019) and 40 Myr
447 ago (Mugerwa et al. 2018).

448 Both *B. tabaci* MEAM1 Bta06115 and Bta06118 proteins consist of a GH17 domain, preceded
449 by a signal peptide, suggesting that they are secreted (Figure 2B). The predicted structures of the
450 Bta06115 and Bta06118 proteins (without the peptide signal) were obtained using AlphaFold2 and
451 aligned with the crystal structure of barley endo-ß-1,3-glucanase and endo-ß-1,3-1,4-glucanase

452 (Varghese et al. 1994). The RMSD values for the alpha carbon atoms ranged from 0.629 Å (for all 255
453 atoms) to 0.767 Å (for all 278 atoms) between the Bta06118 and barley endo-ß-1,3-1,4-glucanase
454 structures and the Bta06115 and barley endo-ß-1,3-glucanase structures, respectively, indicating a high
455 degree of structural similarity (Supplementary Figure 2). Accordingly, the predicted structures of
456 Bta06115 and Bta06118, along with the crystal structures of barley endo-ß-1,3-glucanase and barley
457 endo-ß-1,3-1,4-glucanase, exhibit the typical $(\alpha/ß)_8$ TIM barrel motif shared by GH17 glucanases
458 (Supplementary Figure 2). This motif consists of an inner crown of eight ß-strands linked by loops to
459 an outer crown of eight α-helices (Varghese et al. 1994; Receveur-Bréchot et al. 2006). The two
460 glutamic acid catalytic residues that constitute the active site of GH17 ß-glucanases (Varghese et al.
461 1994) are conserved in the predicted structures of Bta06115 and Bta06118, suggesting the functionality
462 of the *B. tabaci* MEAM1 GH17 proteins (Figure 2C).

463 Bta06115 was heterologously expressed in *P. pastoris* and purified to homogeneity. To
464 investigate substrate preferences of this putative glucanase, we first assayed Bta06115 on cellohexaose
465 (G6; β-1,4-gluco-oligosaccharides) and laminarihexaose (L6; β-1,3-gluco-oligosaccharides) as
466 substrates (Figure 3A). Enzyme assays revealed that Bta06115 was able to efficiently cleave L6 into L2
467 and glucose, suggesting exo-type activity. However, very weak activity was detected on G6. We then
468 tested the enzyme on two different mixed-linked β-glucans (i.e. lichenan and barley glucan), displaying
469 a 1,3/1,4 mean ratio of 1:2 and 1:3, respectively (Figure 3B). Significant activity was detected on
470 lichenan but no activity was detected on barley β-glucan. Therefore, our results indicate a typical
471 lichenase activity for Bta06115 with a significant cleavage of mixed-linked β-1,3/β-1,4 glucans, while
472 no significant cleavage was observed on barley β-glucan, that displays a high ratio of β-1,4-linkages,
473 cellulose or cello-oligosaccharides. Overall, we conclude that the GH17 Bta06115 is a functional β-1,3-
474 glucanase.

475 Accordingly, comparison of the specific residues that determine substrate specificity confirms
476 that Bta06115 and Bta06118 are ß-1,3-glucanases (Figure 2C). In barley endo-ß-1,3-1,4-glucanase,
477 residues Q129 and F135 block the catalytic cleft and prevent the ß-1,3-glucans from binding, while the
478 short T210 leaves sufficient sufficient space for mixed β-1,3-1,4-glucans (Varghese et al. 1994).
479 Residues F130 and S136, which replace Q129 and F135 in barley endo-ß-1,3-glucanase, provide space

480 to accommodate ß-1,3-glucans (Varghese et al. 1994), as would M163 and T169 in Bta06115 and M177
481 and S183 in Bta06118 (Figure 2C). In addition, the replacement of T210 by M209 in barley endo-ß-
482 1,3-glucanase represents a barrier to a mixed β-1,3-1,4-glucan with a high ratio of β-1,4-linkages
483 (Varghese et al. 1994). Residues Q244 and Q260 would form a similar barrier in Bta06115 and
484 Bta06118 respectively (Figure 2C).

485 *A functional GH152 ß-1,3-1,4-glucanase was acquired by HGT from a plant donor in B. tabaci*

486 The *B. tabaci* MEAM1 Bta13961 protein shares similarities with plant ß-glucanases belonging
487 to the GH152 family (Table 2 and Supplementary Table 2). The GH152 family is much less described
488 than the GH17 family (http://www.cazy.org/; Drula et al. 2022), with only one biochemically
489 characterized enzyme from a filamentous fungi (Sakamoto et al. 2006). In our Diamond homology
490 search, homologs of the *B. tabaci* MEAM1 Bta13961 protein were identified in *B. tabaci* MED and
491 SSA-ECA, but not in any organism other than plants (Supplementary Table 2 and Figure 4A). The
492 monophyletic grouping of *B. tabaci* sequences was robust, suggesting horizontal acquisition from a
493 plant donor prior to the divergence of the three *B. tabaci* cryptic species (Figure 4A). The *B. tabaci*
494 MEAM1 Bta13961 was surrounded by non-HGT genes (Supplementary Table 2), and synteny was quite
495 conserved in the genomic region of the Bta13961 gene in *B. tabaci* SSA (Figure 4B), ruling out the
496 possibility of contamination. The Bta13961 protein shares similarities to thaumatin-like (TL) proteins
497 found in plants (Supplementary Table 2). It consists of a signal peptide followed by a thaumatin family
498 domain that covers most, if not all, of the mature peptide (Figure 4C), typical of TL proteins (De Jesús-
499 Pires et al. 2020). Commonly found in plants, TL proteins include pathogenesis-related proteins
500 involved in plant defense against pathogens (De Jesús-Pires et al. 2020). One mechanism of action of
501 some of these TL proteins in plant defense would be their demonstrated endo-ß-1,3-glucanase activity,
502 which would degrade the cell walls of pathogenic fungi (Grenier et al. 1999). TL proteins have also
503 been identified in animals, although no evidence of HGT has yet been demonstrated, and in fungi,
504 including a protein from *Lentinula edodes* with demonstrated ß-1,3-glucanase activity (Sakamoto et al.
505 2006) and classified as the only characterized member of the GH152 family (http://www.cazy.org/;
506 Drula et al. 2022).

507 The predicted structure of the Bta13961 protein (without the peptide signal) was obtained using
508 AlphaFold2 and aligned with the crystal structure of the TL protein NP24-I from tomato (Ghosh and
509 Chakrabarti 2008). The RMSD values for the alpha carbon atoms was 0.693 Å (for all 157 atoms),
510 indicating a high degree of structural similarity (Figure 5A). The REDDD motif that form the acidic
511 cleft responsible for the β-1,3-glucanase activity of TL proteins and that contain the two presumed
512 catalytic residues (De Jesús-Pires et al. 2020) were conserved in the Bta13961 structure, suggesting the
513 functionality of the *B. tabaci* MEAM1 Bta13961 protein (Figure 5A).

514 To investigate the substrate specificity of Bta13961, we used the same approach as for the GH17
515 Bta06115. After successful expression in *P. pastoris* and purification, we first demonstrated that the
516 enzyme was active on both β-1,3 and β-1,4 oligosaccharides using G6 or L6 as substrates (Figure 5B).
517 The activity was significant but in both cases, the substrates were not fully consumed, meaning that
518 they are not the preferred substrates. Enzyme activity was much more significant on β-1,3/β-1,4 glucans,
519 Bta13961 being able to cleave both mixed-linked β-glucans (barley glucan and lichenan) with similar
520 efficiency (Figure 5C). These results clearly indicate a different behavior of the GH152 Bta13961
521 compared to the GH17 Bta06115. Indeed, the 1,3/1,4 mean ratio in β-glucans did not impact enzyme
522 activity and the products released (most probably β-1,3/β-1,4 oligosaccharides) were much longer as
523 compared to the products released by Bta06115. Overall, these results indicated that the GH152
524 Bta13961 is a functional endo-β-1,3-1,4-glucanase able to cleave both β-1,3 and β-1,4 linkages.
525 Interestingly, only β-1,3-glucanase activity has been described for the sole GH152 biochemically
526 characterized to date (http://www.cazy.org/; Drula et al. 2022).

527 *Two CE8 candidate pectin methylesterases were horizontally transferred from a plant donor*
528 *in B. tabaci*

529 The *B. tabaci* MEAM1 Bta11221 and Bta11222 genes encode proteins with similarities to
530 pectin methylesterases (PMEs) found in plants (Supplementary Table 2). PMEs are enzymes of the CE8
531 family of CAZymes that catalyze the demethoxylation of pectin, a major component of the plant cell
532 wall (Pelloux et al. 2007; Jolie et al. 2010). PMEs are ubiquitous enzymes in plants, but are also found

533 in bacterial and fungal phytopathogens with a role in breaking down the plant cell wall to allow
534 infection. PMEs have also been identified in some arthropods (Pauchet et al. 2010; Evangelista et al.
535 2015; Antony et al. 2017; Faddeeva-Vakhrusheva et al. 2017) and in a plant parasitic nematode (Vicente
536 et al. 2019), and were likely horizontally acquired from bacteria.

537 Homologs of the *B. tabaci* MEAM1 Bta11221 and Bta1122 proteins were identified in *B. tabaci*
538 MED and SSA-ECA, but not in any organism other than plants (Supplementary Table 2 and Figure
539 6A). The robust monophyletic grouping of all *B. tabaci* sequences suggests horizontal acquisition from
540 a plant donor prior to the divergence of the three *B. tabaci* cryptic species (Supplementary Table 2 and
541 Figure 6A). The *B. tabaci* MEAM1 Bta11221 and Bta1122 genes were surrounded by genes conserved
542 in other arthropods and animals, which were therefore not horizontally acquired (Supplementary Table
543 2). Furthermore, synteny was conserved in the genomic region of the Bta11221 and Bta1122 genes in
544 *B. tabaci* SSA (Figure 6B), ruling out the possibility of contamination. The two *B. tabaci* PME genes
545 are consecutive in *B. tabaci* MEAM1 and SSA-ECA genomes, suggesting that the HGT event was
546 followed by a tandem duplication before the divergence of the two cryptic species (Figure 6B). Another
547 possibility is that the two genes were organized in tandem in the plant donor and were transferred in the
548 same HGT event. However, this is not consistent with the bootstrap-supported grouping of *B. tabaci*
549 PME proteins into a single monophyletic group (Figure 6A). Both *B. tabaci* MEAM1 Bta11221 and
550 Bta1122 proteins consist of a pectinesterase catalytic domain, preceded by one or two pectinesterase
551 inhibitory domains and, in the case of Bta11222, a signal peptide (Figure 6C). This domain organization,
552 in which the active part of the protein is later cleaved from the N-terminal inhibitory region, is typical
553 for group 2 PMEs found exclusively in plants (Pelloux et al. 2007; Jolie et al. 2010). This is consistent
554 with the hypothesis of an HGT event from a plant donor.

555 The predicted structures of the PME active part of the Bta11221 and Bta1122 proteins were
556 obtained using AlphaFold2 and aligned with the crystal structure of carrot PME, the first to be solved
557 in plants (Johansson et al. 2002). The resulting RMSD values calculated for the alpha carbon atoms
558 were 0.613 Å (for 285 out of 297 atoms) and 0.529 Å (for 171 out of 210 atoms), respectively, indicating
559 a high level of structural similarity (Supplementary Figure 3). The five key residues that were identified
560 in the active site of carrot PME (Johansson et al. 2002; Jolie et al. 2010) are conserved in the predicted

561 structure of Bta11221 and Bta11222 (Supplementary Figure 3). These observations suggest that the
562 Bta11221 and Bta11222 proteins are functional PMEs.

563 *An expansin-related protein was acquired horizontally from a plant donor in B. tabaci*

564 Expansins are enigmatic proteins that promote loosening and extension of the plant cell wall,
565 although they lack enzymatic activity (Cosgrove 2015). Canonical expansins are characterized by the
566 presence of two domains, an N-terminal six-stranded double-psi beta-barrel (DPBB), which is
567 structurally related to GH45, and a C-terminal CBM63. The *B. tabaci* MEAM1 Bta15239 gene encodes
568 a protein with similarities to GH45 endoglucanases (EG45-like domain-containing proteins) from plants
569 (Supplementary Table 2). These are expansin-related proteins that contain the N-terminal expansin-like
570 DPBB domain, but not the C-terminal CBM63 domain found in canonical expansins (Cosgrove 2015).
571 Consistently, the *B. tabaci* MEAM1 Bta15239 protein consists of a single DPBB domain preceded by
572 a signal peptide (Figure 7A). Homologs of the *B. tabaci* MEAM1 Bta15239 protein were identified in
573 the *B. tabaci* MED and SSA-ECA cryptic species, but not in any organism other than plants
574 (Supplementary Table 2 and Figure 7B). The *B. tabaci* MEAM1 Bta15239 gene was surrounded by
575 non-HGT genes (Supplementary Table 2), and synteny was conserved in the genomic region of the
576 Bta15239 gene in *B. tabaci* SSA (Figure 7C), ruling out the possibility of contamination.

577 Expansin-like and expansin-related proteins are present in a large number of plant pathogens
578 among bacteria and fungi with reported roles in plant colonization and virulence (Georgelis et al. 2015;
579 Cosgrove 2017). Evolutionary analyses suggest that these microbial proteins originated from multiple
580 horizontal gene transfers from plants as well as between microbes (Georgelis et al. 2015; Cosgrove
581 2017). In metazoans, expansin-like and expansin-related proteins, probably of bacterial origin, have
582 been described as effectors of parasitism in several plant parasitic nematodes (Danchin et al. 2010).
583 Expansins are thought to act by loosening the interactions between plant cell wall components to
584 facilitate the action of various enzymes also secreted by the nematodes (Qin et al. 2004). However,
585 there is no evidence of cell wall loosening activity for single domain DPBB proteins, which are likely
586 to have a function distinct from that of canonical expansins (Cosgrove 2015).

587 *Much less horizontally acquired CAZymes were found in T. vaporariorum compared to B.*
588 *tabaci*

589 Next, we investigated whether HGT events have made similar contributions to the biology and
590 genome evolution of *T. vaporariorum,* another species in the Aleyrodinae subfamily in the Hemiptera
591 order, which is separated from *B. tabaci* by approximately 80 to 90 Myr (Misof et al. 2014). In their
592 study, Gilbert and Maumus investigated whether some homologs of the *B. tabaci* MEAM1 horizontally
593 acquired genes from plants were conserved in *T. vaporariorum* (Gilbert and Maumus 2022). However,
594 they did not perform a comprehensive search for horizontally acquired genes in the latter. We used AvP
595 to identify HGT candidates in *T. vaporariorum* and compared the results with those obtained in *B.*
596 *tabaci*. Of the 18,725 predicted proteins in *T. vaporariorum*, 182 proteins were identified as likely HGT
597 candidates using AvP (Table 1 and Supplementary Table 6). Of these, 75 candidates specific to
598 Aleyrodinae have been validated with potential donors from bacteria, fungi, or plants, representing less
599 than one-third of the number of validated HGT candidates for *B. tabaci*. The main difference between
600 the two species is in the number of HGT candidates from fungi (8 for *T. vaporariorum* compared to 88
601 for *B. tabaci*) or plants (5 for *T. vaporariorum* compared to 53 for *B. tabaci*) potential donors (Table 1
602 and Supplementary Table 6).

603 The 75 HGT candidates that were validated in *T. vaporariorum* correspond to at least 32 distinct
604 HGT events (Supplementary Table 6). Of these events, 23 showed clustering between sequences of *B.*
605 *tabaci* and *T. vaporariorum*, as determined by Orthofinder. For each of these 23 events, we combined
606 the groups defined by AvP for both *B. tabaci* and *T. vaporariorum*. Phylogenies were then inferred
607 according to the procedure described in the Material and Methods section. The sequences of *B. tabaci*
608 and *T. vaporariorum* formed robust monophyletic clades in 11 of the combined groups, suggesting HGT
609 events that occurred before the two species diverged (Supplementary Figure 4). One of these groups
610 includes Bta02634, classified as a member of the PL1_4 pectin lyase CAZyme family, and Tv_05105-
611 RA, identified as its ortholog by Orthofinder (Table 2). The origin of these genes would correspond to
612 an HGT event in the ancestor of Aleyrodinae, although it could not not be clearly determined from the
613 phylogeny whether the donor belongs to bacteria or fungi (Supplementary Figure 1).

614 Sequences of *B. tabaci* and *T. vaporariorum* formed a monophyletic clade in three of the
615 remaining 12 combined groups, albeit with lower bootstrap support values (Supplementary Table 7). In
616 the remaining nine combined groups, there was no monophyletic clade grouping both *B. tabaci* and *T.*
617 *vaporariorum* sequences (Supplementary Figure 5). Alternative topology tests for monophyly of
618 Aleyrodinae sequences proved inconclusive for seven of them (Supplementary Table 7). Therefore, it
619 is difficult to conclude with confidence that there is a single ancestral HGT event for each of these 11
620 combined groups. One of these groups involves Bta13961, classified as a member of the GH152
621 CAZyme family acquired from plants and demonstrated to be functional as a glucanase (see above),
622 and Tv_15928-RA, identified as its ortholog through Orthofinder. The two sequences shared 42.2%
623 identity and 53.5% similarity at the protein level. Their predicted structures were aligned with an RMSD
624 of 0.586 Å, indicating structural similarity (Figure 8A and Supplementary Figure 1). Despite their
625 similarity, sequences of GH152 of *B. tabaci* and *T. vaporariorum* were found in two different clades in
626 the phylogeny obtained for the combined group (Figure 8B). This may suggest that they originated from
627 two distinct HGT events, but the alternative topology test with a constrained topology supporting the
628 monophyly of Aleyrodinae sequences was not significantly less likely (Supplementary Table 7).
629 Interestingly, the group formed by Tv_15928-RA and two plant sequences included a sequence from *T.*
630 *palmi*, which belongs to the Thysanoptera order, an outgroup of Hemiptera (Figure 8B). This candidate
631 case of HGT in thrips will be further investigated in the next section.

632 Among the 32 HGT events in *T. vaporariorum,* all nine for which no homologs were found in
633 *B. tabaci* have a potential bacterial origin (Supplementary Table 6). Six of these HGT events involve
634 genes that encode proteins from the AAA-ATPase-like domain-containing protein family. As
635 previously mentioned, various HGT events that involve this large family of ATPases have already been
636 identified in *B. tabaci*, and some of them are common to *T. vaporariorum* (Supplementary Table 2).
637 The six HGT events that are unique to *T. vaporariorum* and that involve AAA-ATPase-like domain-
638 containing proteins may correspond to new, independent HGT events of other members of this protein
639 family. Among the three other HGT events unique to *T. vaporariorum*, one concerns the Tv_04079-RA
640 gene, which codes for a CAZyme belonging to the α-glucosidase GH13 family (Supplementary Table
641 6). This *T. vaporariorum*-specific CAZyme is not related to the two *B. tabaci* GH13 CAZymes of the

642 subfamily 17, which have been shown to catalyze the detoxification of glucosinolates produced by
643 plants against herbivores (Malka et al. 2020), but were not identified as HGT candidates. Unfortunately,
644 the sequence of the *T. vaporariorum* GH13 CAZyme acquired horizontally from bacteria is only partial
645 and it is not known to which subfamily it belongs. Overall, only three of the 363 predicted *T. vaporariorum*
646 CAZymes (Supplementary Table 5), one PL1_4, one GH152 and one GH13, would have been acquired by
647 HGT, compared to 25 (out of 433) in *B. tabaci* (Supplementary Tables 2 and 5). This is consistent with the
648 GO analysis, which showed no significant enrichment of GO terms related to carbohydrate metabolism and
649 plant cell wall degradation among HGT candidates for *T. vaporariorum* (Supplementary Table 7).

650 *Few HGT events in piercing-sucking insects other than Aleyrodinae, but involving CAZymes*
651 *with possible activity on plant carbohydrates*

652 Since our results indicate that a member of the GH152 family has been acquired from plants
653 not only in the Aleyrodinae *B. tabaci* and *T. vaporariorum*, but also in *T. palmi* (see above), we used
654 the AvP-based approach to identify potential HGT candidates in *T. palmi* and its related species *F.*
655 *occidentalis*. Both species are cell-content feeding members of the Thripinae subfamily in the
656 Thysanoptera order, which is the closest outgroup to Hemiptera, although the divergence between these
657 two orders probably occurred more than 300 Myr ago (Misof et al. 2014). The number of HGT
658 candidates specific to Thysanoptera found using AvP was low in *F. occidentalis* and *T. palmi* compared
659 to *B. tabaci* and *T. vaporariorum* (Table 1). Only 23 candidates were validated for *F. occidentalis* and
660 only 20 for *T. palmi*. The validated HGT candidates correspond to at least eight HGT events in both *F.*
661 *occidentalis* and *T. palmi*, seven of which are shared by both species (Supplementary Tables 8 and 9).
662 Among the eight HGT events in *F. occidentalis*, we found the three that have been described in the
663 literature with bacteria as potential donors (Rotenberg et al. 2020), further suggesting that our approach
664 is reliable.

665 The analysis of overrepresented GO terms among HGT candidates in *F. occidentalis* and *T.*
666 *palmi* showed a significant enrichment in GO terms related to carbohydrate metabolism and plant cell
667 wall degradation (Supplementary Table 10). These GO terms include hydrolase activity, hydrolyzing
668 O-glycosyl compounds (GO:0004553), cellulase activity (GO:0008810), or carbohydrate metabolic

669 process (GO:0005975). Accordingly, five out of eight HGT events in *F. occidentalis* and six out of
670 eight in *T. palmi* correspond to CAZymes that are predicted to originate from bacteria, fungi, and plants
671 (Table 3, Supplementary Table 5 and Supplementary Figure 5).

672 One of these HGT events corresponding to CAZymes is unique to *T. palmi* and involves a
673 protein of plant origin that contains four carbohydrate-binding modules of the CBM43 family but no
674 associated catalytic module, suggesting that the protein does not directly digest carbohydrates (Table 3
675 and Supplementary Figure 6). The other HGT events corresponding to CAZymes are shared by the two
676 Thripinae species and concern the GH152 family, the GH32 family, subfamily 8 of the GH5 family
677 (GH5_8), the GH45 family, and subfamily 4 of the PL1 family (PL1_4) (Table 3 and Supplementary
678 Tables 8 and 9).

679 Our results confirm that members of the GH152 family of endoglucanases have been acquired
680 from plants not only in the Aleyrodinae *B. tabaci* and *F. occidentalis* (see above), but also in the
681 Thripinae *T. palmi* and *F. occidentalis* (Table 3). In the phylogeny performed on the AvP-defined
682 Thripinae group, the unique sequence found in *F. occidentalis* and the two sequences found in *T. palmi*
683 form a highly-supported monophyletic group, suggesting that they originated from a single HGT event
684 (Supplementary Figure 6). In the phylogeny performed after combining the AvP-defined Aleyrodinae
685 and Thripinae groups, the Aleyrodinae and Thripinae sequences were not clearly separated into distinct
686 clades (Supplementary Figure 6). In addition, the alternative topology constraining the grouping of
687 Aleyrodinae and Thripinae sequences into a single clade was not significantly less likely. Thus, the
688 hypothesis of a single HGT event in the common ancestor of both Aleyrodinae and Thripinae, rather
689 than two independent HGT events, could not be rejected on the basis of the phylogeny alone. However,
690 the hypothesis of a single HGT event would require numerous subsequent loss events. Indeed, the last
691 common ancestor of both Hemiptera and Thysanoptera dates back more than 300 Myr (Misof et al.
692 2014), and no GH152 is found in the numerous other species from these orders (particularly Hemiptera)
693 represented at the sequence level in public libraries. Our results thus suggest a convergent acquisition
694 of GH152 family glucanases in the Aleyrodinae *B. tabaci* and *T. vaporariorum* on the one hand, and
695 the Thripinae *F. occidentalis* and *T. palmi* on the other.

696 Two sequences, which are predicted to have a bacterial origin and are found in both *F.*
697 *occidentalis* and *T. palmi,* belong to the GH32 family. This family includes several horizontally-
698 acquired genes in various metazoans that are involved in plant carbohydrate assimilation, including the
699 metabolism of host-derived sucrose (Nakabachi 2015; Danchin et al. 2016; Wybouw et al. 2016; Dai et
700 al. 2021). The four *F. occidentalis* and *T. palmi* sequences did not form a monophyletic group,
701 suggesting that they resulted from two distinct HGT events (Supplementary Figure 6). However, the
702 alternative topology test using a constrained topology supporting the monophyly of these four sequences
703 was not significantly less likely, and the hypothesis of a single event cannot be rejected. It is worth
704 noting that members of the GH32 family have also been identified as HGT candidates in *B. tabaci* (see
705 above). However, they form a completely distinct clade in a phylogeny performed after combining the
706 AvP-defined *B. tabaci* and Thripinae groups (Supplementary Figure 6), with different bacteria as donors
707 for *B. tabaci* (phylum Pseudomonodata) and for the Thripinae (phylum Actinomycetota). In addition,
708 the alternative topology constraining all Thripinae and *B. tabaci* sequences into one clade was
709 significantly less likely (Supplementary Figure 6). This, together with the absence of homologous
710 sequences in other members of the Thysanoptera and Hemiptera, suggests two independent and
711 convergent acquisition events.

712 Three sequences found in *F. occidentalis* and two in *T. palmi*, all predicted to originate from
713 bacteria, belong to the GH5_8 family of endomannanases (Table 3). All five sequences form a highly-
714 supported monophyletic group, indicating that they originated from a single HGT event (Supplementary
715 Figure 6). GH5_8 enzymes, which degrade mannan, have reportedly been acquired by several other
716 arthropods through HGT, possibly playing a role in plant cell wall degradation (Acuña et al. 2012; Vega
717 et al. 2015; Shin et al. 2021).

718 Three sequences found in *F. occidentalis* and two in *T. palmi*, all predicted to originate from
719 fungi, belong to the G45 family of endoglucanases capable of hydrolyzing cellulose (Table 3). All five
720 sequences constitute a monophyletic group supported by bootstrap, suggesting a single HGT event
721 origin (Supplementary Figure 6). Sequences from other arthropods, as well as from nematodes and
722 rotifers, appear in the phylogeny. Although they also share similarities with GH45 proteins from fungi,
723 they form distinct clades in the phylogeny (Supplementary Figure 6). This was not unexpected, as it has

724 been shown that GH45 enzymes, with cellulase activity characterized in some cases, were probably
725 acquired by HGT independently in various phytophagous and phytoparasitic organisms (Palomares-
726 Rius et al. 2014; Szydlowski et al. 2015; Busch et al. 2019).

727 The last HGT event corresponding to a CAZyme and shared by the two Thripinae species
728 concerns 10 sequences found in *F. occidentalis* and nine in *T. palmi* belonging to the PL1_4 family of
729 pectin lyases (Table 3). All sequences form a highly-supported monophyletic group, indicating that they
730 originated from a single HGT event, although we could not clearly determine from the phylogeny
731 whether the donor belonged to bacteria or fungi (Supplementary Figure 6). It is worth noting that a
732 member of this family has been identified as an HGT candidate in *B. tabaci* and *T. vaporariorum* (see
733 above), and that sequences from two other insects, belonging to Hemiptera and Hymenoptera, are found
734 in the phylogeny (Supplementary Figure 6). In the phylogeny performed after combining the AvP-
735 defined Aleyrodinae and Thripinae groups, the Hemiptera and Thysanoptera sequences were not
736 separated into distinct clades, in contrast to the Hymenoptera sequences (Supplementary Figure 6). The
737 alternative topology, which constrains all metazoan sequences to be grouped in a single clade, was not
738 significantly less likely (Supplementary Figure 1). The hypothesis of a single ancestral HGT event could
739 therefore not be rejected, although it would require numerous subsequent loss events. Indeed, the last
740 common ancestor of Hemiptera, Hymenoptera, and Thysanoptera dates back approximately to 360 Myr
741 (Misof et al. 2014), and most of the descendant species lack this gene while being represented at the
742 sequence level in public libraries.

743

744 **Conclusion**

745 In the present study, we used the AvP software package (Koutsovoulos et al. 2022) followed by
746 manual validation to perform a comprehensive and accurate detection of HGT candidates from non-
747 metazoan donors in the related phloem-feeding insects *B. tabaci* MEAM1 and *T. vaporariorum*,
748 members of the subfamily Aleyrodinae of the order Hemiptera. With a total of 255 HGT candidates
749 (corresponding to 136 distinct HGT events), our results confirm findings from previous studies
750 indicating that a large number of genes have been acquired by HGT from bacterial, fungal, but also
751 plant donors in *B. tabaci* (Chen et al. 2016; Gilbert and Maumus 2022; Li et al. 2022). In contrast, the

752 number of HGT candidates in the related species *T. vaporariorum* was substantially lower, with 75
753 HGT candidates corresponding to a total of 32 HGT events, of which only three represented gene
754 acquisition from plants (compared to 27 for *B. tabaci*). The majority of HGT events found in *T.*
755 *vaporariorum* appear to have occurred in the common ancestor of the Aleyrodinae, whereas the majority
756 of HGT events found in *B. tabaci* MEAM1, including the plant-acquired BtPMaT1 gene involved in
757 detoxification of plant toxins (Xia et al. 2021), are more recent and occurred in the common ancestor
758 of the three cryptic species of *B. tabaci*. An even lower number of HGT candidates was obtained from
759 a similar automated HGT detection and manual validation analysis performed on the plant cell-feeding
760 insects, *F. occidentalis* (23 HGT candidates corresponding to eight HGT events) and *T. palmi* (20 HGT
761 candidates corresponding to eight HGT events), members of the subfamily Thripinae of the order
762 Thysanoptera. Most of the HGT events are shared between *F. occidentalis* and *T. palmi* and probably
763 occurred in the common ancestor of the Thripinae, in contrast to the Aleyrodinae, for which most of the
764 HGT events are unique to *B. tabaci*. The much higher number of HGT candidates in *B. tabaci* compared
765 to the three other species analyzed is consistent with the results of a recent study of 218 insect species,
766 in which *B. tabaci* had by far the highest number of HGT-acquired genes (Li et al. 2022). An open
767 question would be to determine the reasons why the number of HGT-acquired genes, especially from
768 plants, is so high in *B. tabaci* compared to *T. vaporariorum*, but also to other piercing-sucking insect
769 species.

770 In this study we focused on the horizontal acquisition of CAZymes, given their significance in
771 phytophagous arthropods (Kirsch et al. 2014; Nakabachi 2015; Wybouw et al. 2016; Husnik and
772 McCutcheon 2018). A total of 25 HGT-acquired CAZymes (corresponding to 14 HGT events) were
773 identified for *B. tabaci*, with potential roles including plant cell wall degradation to facilitate stylet
774 penetration into a phloem sieve element or assimilation of carbohydrates found in phloem. In contrast,
775 only three CAZymes acquired by HGT were found in the related *T. vaporariorum*, one of which would
776 correspond to a HGT event unique to this species. This suggests that the HGT acquisition of CAZymes
777 has had a less significant impact on the biology of *T. vaporariorum*, compared to *B. tabaci*. The 14
778 HGT events corresponding to CAZymes identified for *B. tabaci* involved not only bacteria or fungi as
779 potential donors, as is typically described, but also plants for as many as seven of them. A functional

780 analysis was conducted on two of the plant-acquired *B. tabaci* CAZymes, belonging to the GH17 and
781 GH152 families. Both CAZymes were found to be functional glucanases, exhibiting different activities.
782 The GH17 enzyme exhibited ß-1,3-glucanase activity, while the GH152 enzyme exhibited ß-1,3-1,4-
783 glucanase activity, capable of cleaving both β-1,3 and β-1,4 linkages. These findings suggest that the
784 GH17 and GH152 *B. tabaci* CAZymes may have distinct functional roles, given their different substrate
785 specificities. One of these enzymes may be involved in counteracting one of the mechanisms by which
786 plants defend themselves against phloem-feeding insects. This involves the occlusion of the punctured
787 sieve element by the deposition of callose, which is composed of β-1,3-glucan (Silva-Sanzana et al.
788 2020; Walker 2022).

789 A member of the GH152 family was also identified as having been acquired from plants in *T.*
790 *vaporariorum*, suggesting a unique HGT event in the ancestor of the Aleyrodinae. However, our results
791 do not exclude the possibility that it originated from a distinct HGT event. It is noteworthy that a HGT-
792 acquired GH152 CAZyme with a plant as a potential donor was identified as well in the Thripinae *F.*
793 *occidentalis* and *T. palmi*. This suggests a convergent acquisition of a GH152 CAZyme in the unrelated
794 Aleyrodinae and Thripinae, indicating the potential importance of this enzyme for piercing-sucking
795 insects. Although the number of HGT events identified for *F. occidentalis* and *T. palmi* was relatively
796 low compared to the Aleyrodinae, the majority of them corresponded to the acquisition of CAZymes,
797 several of which had predicted functions important in the plant-insect interaction. This indicates that
798 HGT-acquired CAZymes have biological significance not only for Aleyrodinae, and particularly *B.*
799 *tabaci*, but also for other piercing-sucking insects, including the Thripinae *F. occidentalis* and *T. palmi*.
800

801 **Acknowledgements**

802 We are grateful to the BIG bioinformatics platform from the PlantBios infrastructure for providing
803 facilities and technical support (ISC PlantBios, https://doi.org/10.15454/qyey-ar89) and to the genotoul
804 bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul,
805 https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. Part of the
806 work described was performed using services provided by the 3PE platform, a member of IBISBA-FR

807 (https://doi.org/10.15454/08BX-VJ91; www.ibisba.fr), the French node of the European research
808 infrastructure, EU-IBISBA (www.ibisba.eu).
809

810 **Data and resource availability**

811 Phylogenetic trees generated by IQ-TREE for validated HGT candidates from potential bacterial,
812 fungal, or viridiplantae donors for *B. tabaci*, *T. vaporariorum*, *F. occidentalis*, and *T. palmi* are available
813 as supplementary datasets online at Data INRAE: https://doi.org/10.57745/6GC9WA
814

815 **Author contributions**

816 Conceived and designed the study: DC, JGB, EGJD
817 Performed experiments: DC, MH, MB, SG, GDK
818 Analyzed the data: DC, ED, MB, SG, JGB, EGJD
819 Contributed analysis tools: CB
820 Wrote the manuscript: DC, EGJD, with contribution from JGB
821 All authors read and approved the final version of the manuscript.
822

823 **References**

824 Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH,
825 Egan AN, Doyle JJ, et al. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive
826 insect pest of coffee. *Proc. Natl. Acad. Sci.* 109:4197–4202.
827 Antony B, Johny J, Aldosari SA, Abdelazim MM. 2017. Identification and expression profiling of
828 novel plant cell wall degrading enzymes from a destructive pest of palm trees,
829 Rhynchophorus ferrugineus. *Insect Mol. Biol.* 26:469–484.
830 Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria.
831 *Nat. Rev. Microbiol.* 20:206–218.
832 Bissaro B, Kodama S, Nishiuchi T, Díaz-Rovira AM, Hage H, Ribeaucourt D, Haon M, Grisel S,
833 Simaan AJ, Beisson F, et al. 2022. Tandem metalloenzymes gate plant cell entry by
834 pathogenic fungi. *Sci. Adv.* 8:eade9982.

835 Buchfink B, Reuter K, Drost H-G. 2021. Sensitive protein alignments at tree-of-life scale using
836 DIAMOND. *Nat. Methods* 18:366–368.
837 Busch A, Danchin EGJ, Pauchet Y. 2019. Functional diversification of horizontally acquired
838 glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles. *BMC Evol. Biol.*
839 19:100.
840 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment
841 trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
842 Chen R, Huangfu L, Lu Y, Fang H, Xu Y, Li P, Zhou Y, Xu C, Huang J, Yang Z. 2021. Adaptive
843 innovation of green plants by horizontal gene transfer. *Biotechnol. Adv.* 46:107671.
844 Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun
845 H, et al. 2016. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest,
846 provides novel insights into virus transmission, host adaptation, and insecticide resistance.
847 *BMC Biol.* 14:110.
848 Chen W, Wosula EN, Hasegawa DK, Casinga C, Shirima RR, Fiaboe KKM, Hanna R, Fosto A,
849 Goergen G, Tamò M, et al. 2019. Genome of the African cassava whitefly Bemisia tabaci and
850 distribution and genetic diversity of cassava-colonizing whiteflies in Africa. *Insect Biochem.*
851 *Mol. Biol.* 110:112–120.
852 Cosgrove DJ. 2015. Plant expansins: diversity and interactions with plant cell walls. *Curr. Opin. Plant*
853 *Biol.* 25:162–172.
854 Cosgrove DJ. 2017. Microbial expansins. *Annu. Rev. Microbiol.* 71:479–497.
855 Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. 2015. Expression of multiple horizontally
856 acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.*
857 16:50.
858 Dai X, Kiuchi T, Zhou Y, Jia S, Xu Y, Katsuma S, Shimada T, Wang H. 2021. Horizontal Gene
859 Transfer and Gene Duplication of β-Fructofuranosidase Confer Lepidopteran Insects
860 Metabolic Benefits. *Mol. Biol. Evol.* 38:2897–2914.
861 Danchin EGJ, Guzeeva EA, Mantelin S, Berepiki A, Jones JT. 2016. Horizontal Gene Transfer from
862 Bacteria Has Enabled the Plant-Parasitic Nematode Globodera pallida to Feed on Host-

863 Derived Sucrose. *Mol. Biol. Evol.* 33:1571–1579.
864 Danchin EGJ, Rosso M-N. 2012. Lateral gene transfers have polished animal genomes: lessons from
865 nematodes. *Front. Cell. Infect. Microbiol.* 2:27.
866 Danchin EGJ, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010.
867 Multiple lateral gene transfers and duplications have promoted plant parasitism ability in
868 nematodes. *Proc. Natl. Acad. Sci. U. S. A.* 107:17651–17656.
869 De Jesús-Pires C, Ferreira-Neto JRC, Pacifico Bezerra-Neto J, Kido EA, De Oliveira Silva RL,
870 Pandolfi V, Wanderley-Nogueira AC, Binneck E, Da Costa AF, Pio-Ribeiro G, et al. 2020.
871 Plant Thaumatin-like Proteins: Function, Evolution and Biotechnological Applications. *Curr.*
872 *Protein Pept. Sci.* 21:36–51.
873 Drezen J-M, Josse T, Bézier A, Gauthier J, Huguet E, Herniou EA. 2017. Impact of Lateral Transfers
874 on the Genomes of Lepidoptera. *Genes* 8:315.
875 Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active
876 enzyme database: functions and literature. *Nucleic Acids Res.* 50:D571–D577.
877 Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons
878 dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.
879 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics.
880 *Genome Biol.* 20:238.
881 Evangelista DE, de Paula FFP, Rodrigues A, Henrique-Silva F. 2015. Pectinases from Sphenophorus
882 levis Vaurie, 1978 (Coleoptera: Curculionidae): putative accessory digestive enzymes. *J.*
883 *Insect Sci. Online* 15:168.
884 Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W,
885 Kampfraath AA, Ellers J, Le Ngoc G, van Gestel CAM, et al. 2017. Coping with living in the
886 soil: the genome of the parthenogenetic springtail Folsomia candida. *BMC Genomics* 18:493.
887 Georgelis N, Nikolaidis N, Cosgrove DJ. 2015. Bacterial expansins and related proteins from the
888 world of microbes. *Appl. Microbiol. Biotechnol.* 99:3807–3823.
889 Ghosh R, Chakrabarti C. 2008. Crystal structure analysis of NP24-I: a thaumatin-like protein. *Planta*
890 228:883–890.

891 Gilbert C, Maumus F. 2022. Multiple Horizontal Acquisitions of Plant Genes in the Whitefly Bemisia
892 tabaci. *Genome Biol. Evol.* 14:evac141.
893 Gong C, Guo Z, Hu Y, Yang Z, Xia J, Yang X, Xie W, Wang S, Wu Q, Ye W, et al. 2023. A
894 Horizontally Transferred Plant Fatty Acid Desaturase Gene Steers Whitefly Reproduction.
895 *Adv. Sci.* 11:2306653.
896 Grenier J, Potvin C, Trudel J, Asselin A. 1999. Some thaumatin-like proteins hydrolyse polymeric β-
897 1,3-glucans. *Plant J.* 19:473–480.
898 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and
899 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML
900 3.0. *Syst. Biol.* 59:307–321.
901 Haegeman A, Jones JT, Danchin EGJ. 2011. Horizontal Gene Transfer in Nematodes: A Catalyst for
902 Plant Parasitism? *Mol. Plant-Microbe Interactions®* 24:879–887.
903 Haon M, Grisel S, Navarro D, Gruet A, Berrin J-G, Bignon C. 2015. Recombinant protein production
904 facility for fungal biomass-degrading enzymes using the yeast Pichia pastoris. *Front.*
905 *Microbiol.* 6:1002.
906 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the
907 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522.
908 Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat.*
909 *Rev. Microbiol.* 16:67–79.
910 Johansson K, El-Ahmad M, Friemann R, Jörnvall H, Markovič O, Eklund H. 2002. Crystal structure
911 of plant pectin methylesterase. *FEBS Lett.* 514:243–249.
912 Jolie RP, Duvetter T, Van Loey AM, Hendrickx ME. 2010. Pectin methylesterase and its
913 proteinaceous inhibitor: a review. *Carbohydr. Res.* 345:2583–2595.
914 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
915 Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification.
916 *Bioinforma. Oxf. Engl.* 30:1236–1240.
917 Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of
918 hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.

919 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast
920 model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.
921 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
922 Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780.
923 Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.*
924 9:605–618.
925 Kirsch R, Gramzow L, Theißen G, Siegfried BD, Ffrench-Constant RH, Heckel DG, Pauchet Y. 2014.
926 Horizontal gene transfer and functional diversification of plant cell wall degrading
927 polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect Biochem. Mol.*
928 *Biol.* 52:33–50.
929 Koutsovoulos GD, Noriot SG, Bailly-Bechet M, Danchin EGJ, Rancurel C. 2022. AvP: A software
930 package for automatic phylogenetic detection of candidate horizontal gene transfers. *PLOS*
931 *Comput. Biol.* 18:e1010686.
932 Lapadula WJ, Mascotti ML, Juri Ayub M. 2020. Whitefly genomes contain ribotoxin coding genes
933 acquired from plants. *Sci. Rep.* 10:15503.
934 Le NG, van Ulsen P, van Spanning R, Brouwer A, van Straalen NM, Roelofs D. 2022. A Functional
935 Carbohydrate Degrading Enzyme Potentially Acquired by Horizontal Gene Transfer in the
936 Genome of the Soil Invertebrate Folsomia candida. *Genes* 13:1402.
937 Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
938 display and annotation. *Nucleic Acids Res.* 49:W293–W296.
939 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
940 nucleotide sequences. *Bioinformatics* 22:1658–1659.
941 Li Y, Liu Z, Liu C, Shi Z, Pang L, Chen C, Chen Y, Pan R, Zhou W, Chen X, et al. 2022. HGT is
942 widespread in insects and contributes to male courtship in lepidopterans. *Cell* 185:2975-
943 2987.e10.
944 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and
945 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring
946 of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38:4647–4654.

947 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R.
948 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
949 Genomic Era. *Mol. Biol. Evol.* 37:1530–1534.
950 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making
951 protein folding accessible to all. *Nat. Methods* 19:679–682.
952 Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T,
953 Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution.
954 *Science* 346:763–767.
955 Mitreva-Dautova M, Roze E, Overmars H, de Graaff L, Schots A, Helder J, Goverse A, Bakker J,
956 Smant G. 2006. A symbiont-independent endo-1,4-β-xylanase from the plant-parasitic
957 nematode Meloidogyne incognita. *Mol. Plant-Microbe Interactions®* 19:521–529.
958 Mugerwa H, Seal S, Wang H-L, Patel MV, Kabaalu R, Omongo CA, Alicai T, Tairo F, Ndunguru J,
959 Sseruwagi P, et al. 2018. African ancestry of New World, Bemisia tabaci-whitefly species.
960 *Sci. Rep.* 8:2734.
961 Nakabachi A. 2015. Horizontal gene transfers in insects. *Curr. Opin. Insect Sci.* 7:24–29.
962 Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. 2011. Emerging Virus Diseases Transmitted by
963 Whiteflies. *Annu. Rev. Phytopathol.* 49:219–248.
964 Padilla-Hurtado B, Flórez-Ramos C, Aguilera-Gálvez C, Medina-Olaya J, Ramírez-Sanjuan A,
965 Rubio-Gómez J, Acuña-Zornosa R. 2012. Cloning and expression of an endo-1,4-β-xylanase
966 from the coffee berry borer, Hypothenemus hampei. *BMC Res. Notes* 5:23.
967 Palomares-Rius JE, Hirooka Y, Tsai IJ, Masuya H, Hino A, Kanzaki N, Jones JT, Kikuchi T. 2014.
968 Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and
969 fungi. *BMC Evol. Biol.* 14:69.
970 Pauchet Y, Heckel DG. 2013. The genome of the mustard leaf beetle encodes two active xylanases
971 originally acquired from bacteria through horizontal gene transfer. *Proc. R. Soc. B Biol. Sci.*
972 280:20131021.
973 Pauchet Y, Kirsch R, Giraud S, Vogel H, Heckel DG. 2014. Identification and characterization of
974 plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid

975 beetle Apriona japonica. *Insect Biochem. Mol. Biol.* 49:1–13.
976 Pauchet Y, Wilkinson P, Chauhan R, ffrench-Constant RH. 2010. Diversity of Beetle Genes Encoding
977 Novel Plant Cell Wall Degrading Enzymes. *PLOS ONE* 5:e15635.
978 Pelloux J, Rustérucci C, Mellerowicz EJ. 2007. New insights into pectin methylesterase structure and
979 function. *Trends Plant Sci.* 12:267–277.
980 Perrot T, Pauly M, Ramírez V. 2022. Emerging Roles of β-Glucanases in Plant Development and
981 Adaptative Responses. *Plants* 11:1119.
982 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for
983 Large Alignments. *PLOS ONE* 5:e9490.
984 Prüfer K, Muetzel B, Do H-H, Weiss G, Khaitovich P, Rahm E, Pääbo S, Lachmann M, Enard W.
985 2007. FUNC: a package for detecting significant associations between gene sets and
986 ontological annotations. *BMC Bioinformatics* 8:41.
987 Qin L, Kudla U, Roze EHA, Goverse A, Popeijus H, Nieuwland J, Overmars H, Jones JT, Schots A,
988 Smant G, et al. 2004. Plant degradation: a nematode expansin acting on plants. *Nature*
989 427:30.
990 Receveur-Bréchot V, Czjzek M, Barre A, Roussel A, Peumans WJ, Van Damme EJM, Rougé P.
991 2006. Crystal structure at 1.45-Å resolution of the major allergen endo-β-1,3-glucanase of
992 banana as a molecular basis for the latex-fruit syndrome. *Proteins Struct. Funct. Bioinforma.*
993 63:235–242.
994 Ren F-R, Sun X, Wang T-Y, Yao Y-L, Huang Y-Z, Zhang X, Luan J-B. 2020. Biotin provisioning by
995 horizontally transferred genes from bacteria confers animal fitness benefits. *ISME J.* 14:2542–
996 2553.
997 Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, Jacobs CGC,
998 Vargas Jentzsch IM, Oliver JE, Poelchau MF, et al. 2020. Genome-enabled insights into the
999 biology of thrips as crop pests. *BMC Biol.* 18:142.
1000 Sakamoto Y, Watanabe H, Nagai M, Nakade K, Takahashi M, Sato T. 2006. Lentinula edodes tlg1
1001 Encodes a Thaumatin-Like Protein That Is Involved in Lentinan Degradation and Fruiting
1002 Body Senescence. *Plant Physiol.* 141:793–801.

1003 Shimodaira H. 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. *Syst. Biol.*
1004 51:492–508.
1005 Shin NR, Pauchet Y. 2023. First evidence of a horizontally-acquired GH-7 cellobiohydrolase from a
1006 longhorned beetle genome. *Arch. Insect Biochem. Physiol.* 114:e22039.
1007 Shin NR, Shin S, Okamura Y, Kirsch R, Lombard V, Svacha P, Denux O, Augustin S, Henrissat B,
1008 McKenna DD, et al. 2021. Larvae of longhorned beetles (Coleoptera; Cerambycidae) have
1009 evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes.
1010 *Syst. Entomol.* 46:784–797.
1011 Silva-Sanzana C, Estevez JM, Blanco-Herrera F. 2020. Influence of cell wall polymers and their
1012 modifying enzymes during plant–aphid interactions. *J. Exp. Bot.* 71:3854–3864.
1013 Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison.
1014 *BMC Bioinformatics* 6:31.
1015 Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories
1016 of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-
1017 feeding insects. *Mol. Biol. Evol.* 31:857–871.
1018 Smant G, Stokkermans JPWG, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ,
1019 Henrissat B, Davis EL, et al. 1998. Endogenous cellulases in animals: Isolation of β-1,4-
1020 endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc. Natl. Acad.*
1021 *Sci.* 95:4906–4911.
1022 Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat. Rev.*
1023 *Genet.* 16:472–482.
1024 Sun BF, Xiao JH, He SM, Liu L, Murphy RW, Huang DW. 2013. Multiple ancient horizontal gene
1025 transfers and duplications in lepidopteran species. *Insect Mol. Biol.* 22:72–87.
1026 Szydlowski L, Boschetti C, Crisp A, Barbosa EGG, Tunnacliffe A. 2015. Multiple horizontally
1027 acquired genes from fungal and prokaryotic donors encode cellulolytic enzymes in the
1028 bdelloid rotifer Adineta ricciae. *Gene* 566:125–137.
1029 Varghese JN, Garrett TP, Colman PM, Chen L, Høj PB, Fincher GB. 1994. Three-dimensional
1030 structures of two plant beta-glucan endohydrolases with distinct substrate specificities. *Proc.*

1031 *Natl. Acad. Sci. U. S. A.* 91:2785–2789.
1032 Vega FE, Brown SM, Chen H, Shen E, Nair MB, Ceja-Navarro JA, Brodie EL, Infante F, Dowd PF,
1033 Pain A. 2015. Draft genome of the most devastating insect pest of coffee worldwide: the
1034 coffee berry borer, Hypothenemus hampei. *Sci. Rep.* 5:12525.
1035 Vicente CSL, Nemchinov LG, Mota M, Eisenback JD, Kamo K, Vieira P. 2019. Identification and
1036 characterization of the first pectin methylesterase gene discovered in the root lesion nematode
1037 Pratylenchus penetrans. *PLOS ONE* 14:e0212540.
1038 Walker GP. 2022. Sieve element occlusion: Interactions with phloem sap-feeding insects. A review. *J.*
1039 *Plant Physiol.* 269:153582.
1040 Wang Y-Z, Ye Y-X, Lu J-B, Wang X, Lu H-B, Zhang Z-L, Ye Z-X, Lu Y-W, Sun Z-T, Chen J-P, et
1041 al. 2023. Horizontally Transferred Salivary Protein Promotes Insect Feeding by Suppressing
1042 Ferredoxin-Mediated Plant Defenses. *Mol. Biol. Evol.* 40:msad221.
1043 Wybouw N, Pauchet Y, Heckel DG, Van Leeuwen T. 2016. Horizontal Gene Transfer Contributes to
1044 the Evolution of Arthropod Herbivory. *Genome Biol. Evol.* 8:1785–1801.
1045 Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, Yang X, Yang F, Wu Q, Xie W, et al. 2021. Whitefly
1046 hijacks a plant detoxification gene that neutralizes plant toxins. *Cell* 184:1693-1705.e17.
1047 Xie W, Chen C, Yang Z, Guo L, Yang X, Wang D, Chen M, Huang J, Wen Y, Zeng Y, et al. 2017.
1048 Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q. *GigaScience* 6:1–7.
1049 Xie W, He C, Fei Z, Zhang Y. 2020. Chromosome-level genome assembly of the greenhouse whitefly
1050 (Trialeurodes vaporariorum Westwood). *Mol. Ecol. Resour.* 20:995–1006.
1051

1052 **Figures**

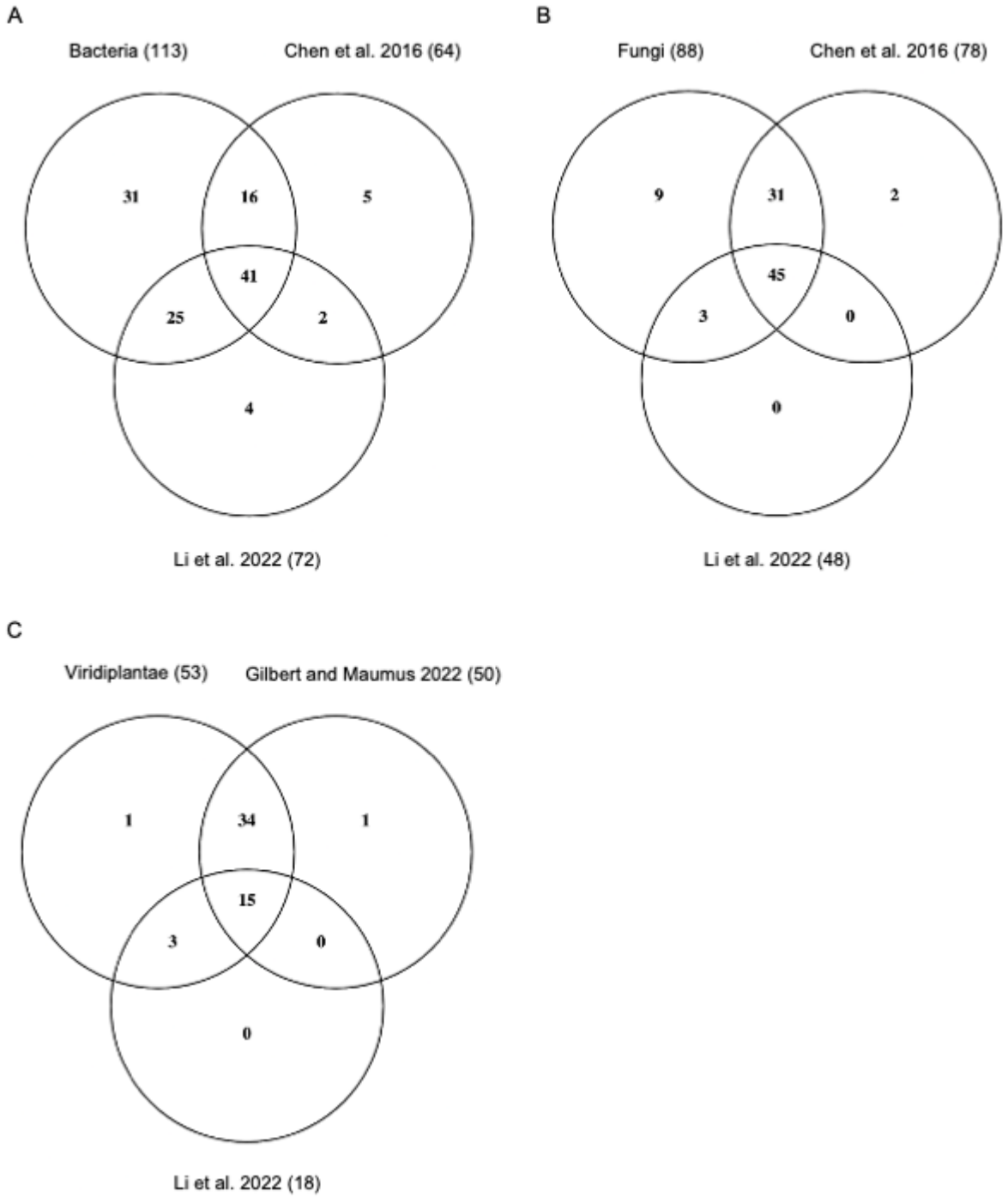
1053 **Figure 1.** Comparison of the number of HGT candidates found for *B. tabaci* MEAM1 with bacteria

1054 (A), fungi (B) or plant (C) as donors with data from the literature.

1055

1056

1057

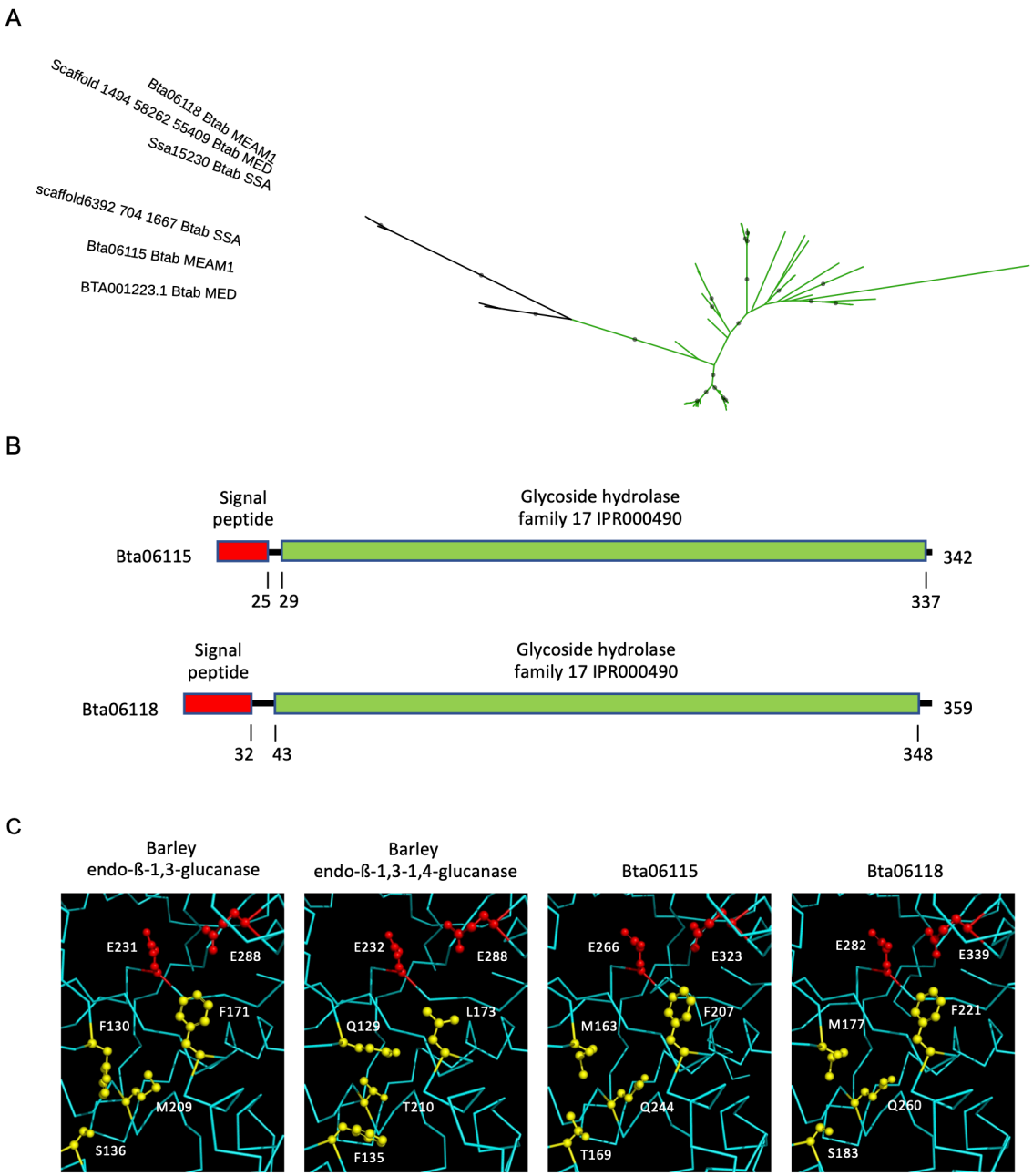


1058 **Figure 2.** Analysis of the GH17 family proteins acquired by HGT in *B. tabaci*. (A) Phylogenetic tree
1059 with *B. tabaci* MEAM1, MED and SSA sequences shown in black. Green branches represent plant
1060 sequences. Gray circles indicate nodes with support values greater than or equal to 80% and 95% for
1061 SH-aLRT and UFboot, respectively. (B) Domain organization of the *B. tabaci* MEAM1 Bta06115 and
1062 Bta06118 proteins. (C) Comparison of the two catalytic glutamate residues (red) and specific residues
1063 that determine substrate specificity (yellow) in the catalytic cleft between the crystal structures of barley
1064 endo-ß-1,3-glucanase and endo-ß-1,3-1,4-glucanase and the predicted structures of *B. tabaci* MEAM1
1065 Bta06115 and Bta06118 proteins.

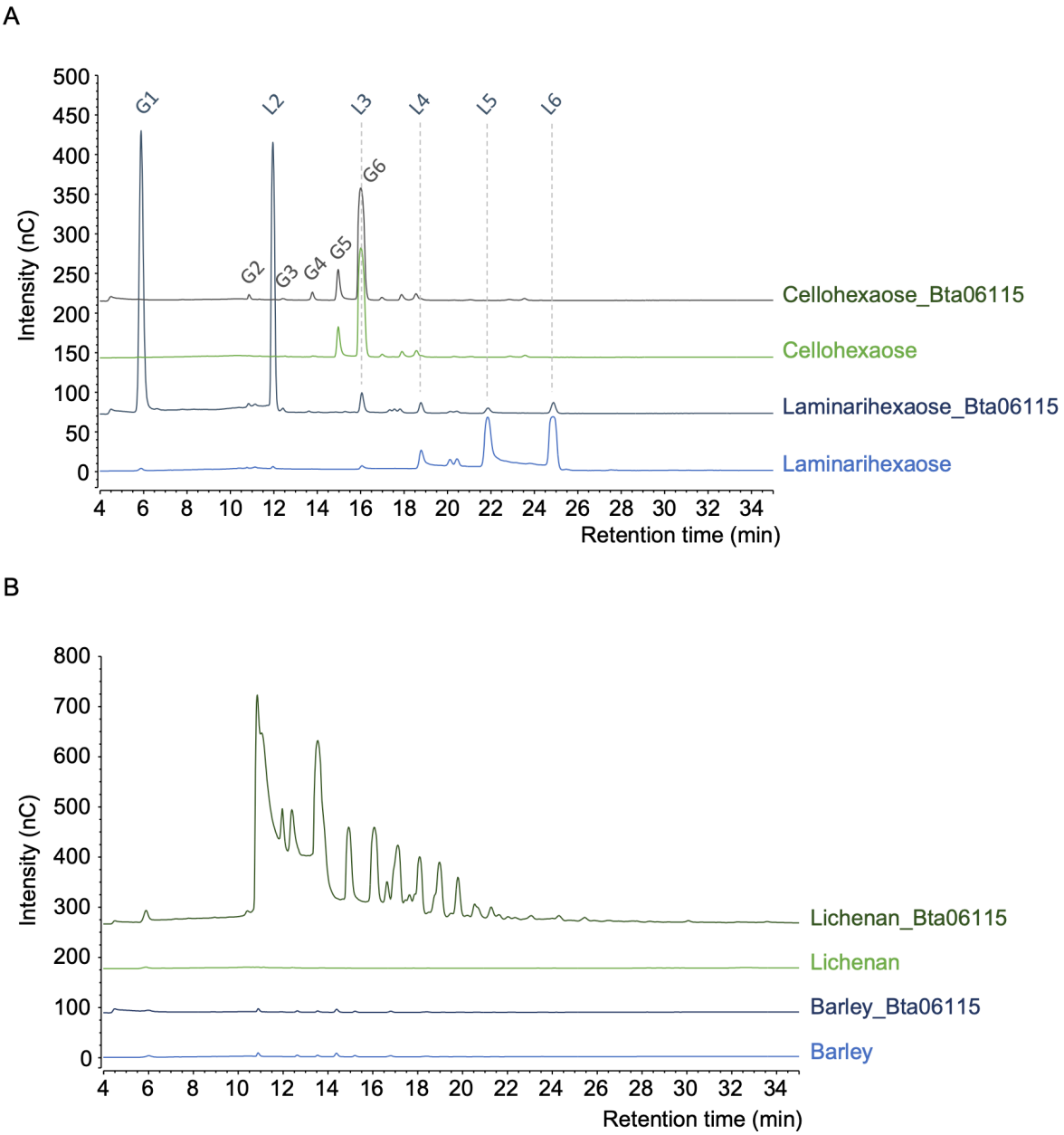
1066

1067

1068

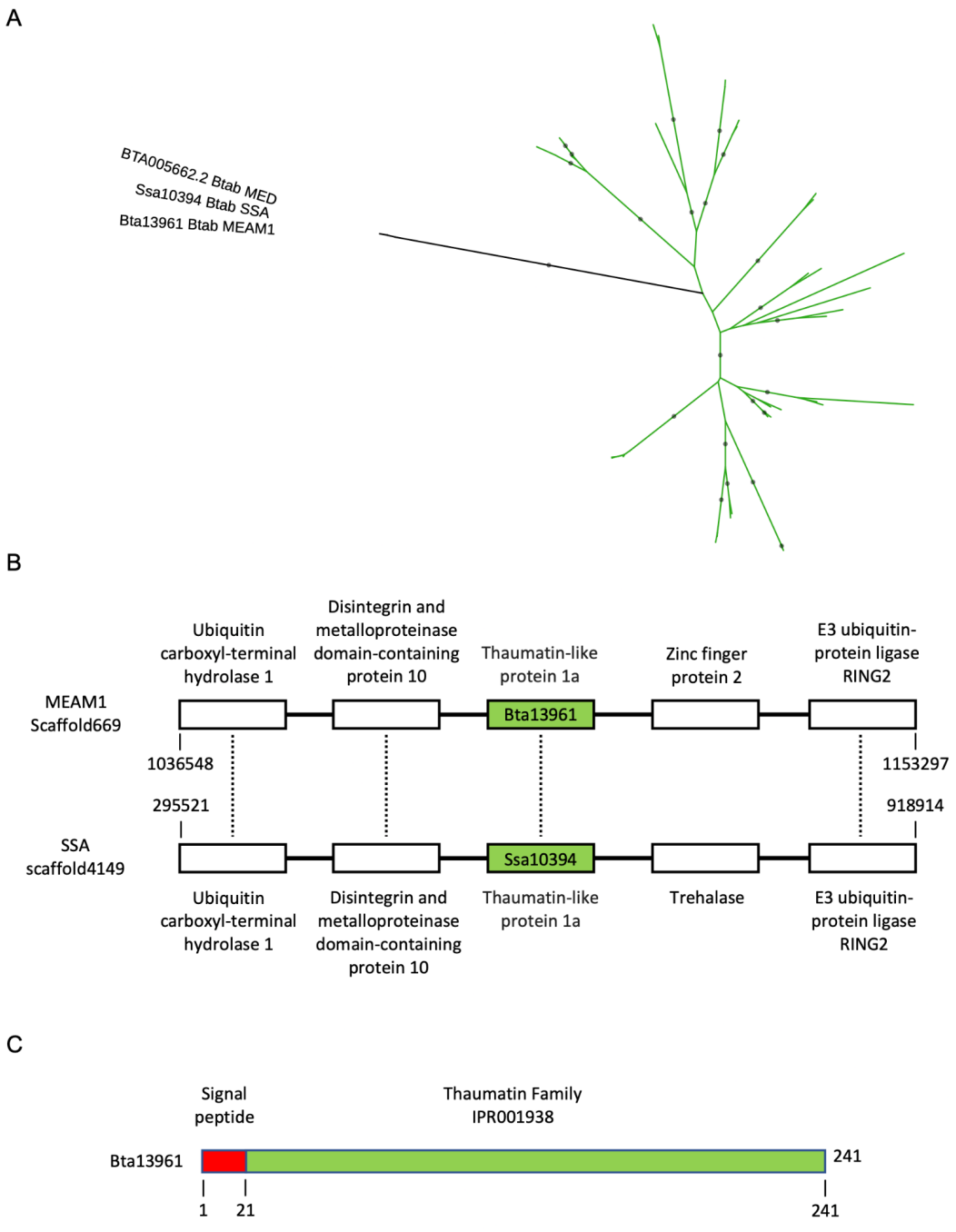


1069 **Figure 3.** Functional analysis of the GH17 family protein Bta06115 acquired by HGT in *B. tabaci*. The
1070 graphs show HPAEC-PAD chromatograms of reaction products released by the GH17 Bta06115 from
1071 cellohexaose (G6) and laminarihexaose (L6) (A) and from barley glucan and lichenan (B).
1072
1073

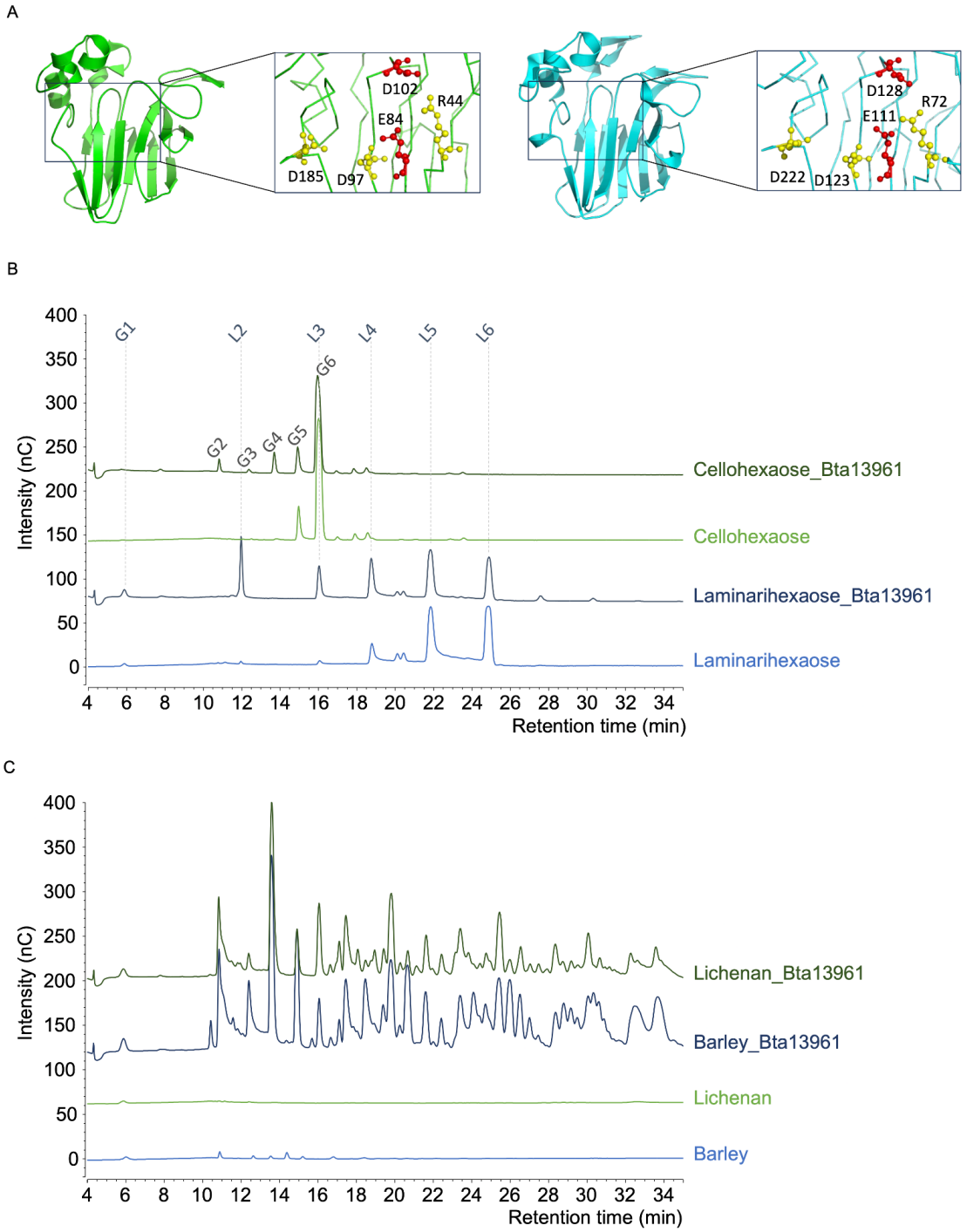


1074 **Figure 4.** Analysis of the GH152 family protein Bta13961 acquired by HGT in *B. tabaci*. (A)
1075 Phylogenetic tree with *B. tabaci* MEAM1, MED and SSA sequences shown in black. Green branches
1076 represent plant sequences. Gray circles indicate nodes with support values greater than or equal to 80%
1077 and 95% for SH-aLRT and UFboot, respectively. (B) Comparative map of the genomic region of the
1078 gene encoding the GH152 family protein in *B. tabaci* MEAM1 (Bta13961) and SSA (Ssa10394) Each
1079 rectangle represents a *B. tabaci* gene. (C) Domain organization of the *B. tabaci* MEAM1 Bta13961
1080 protein.

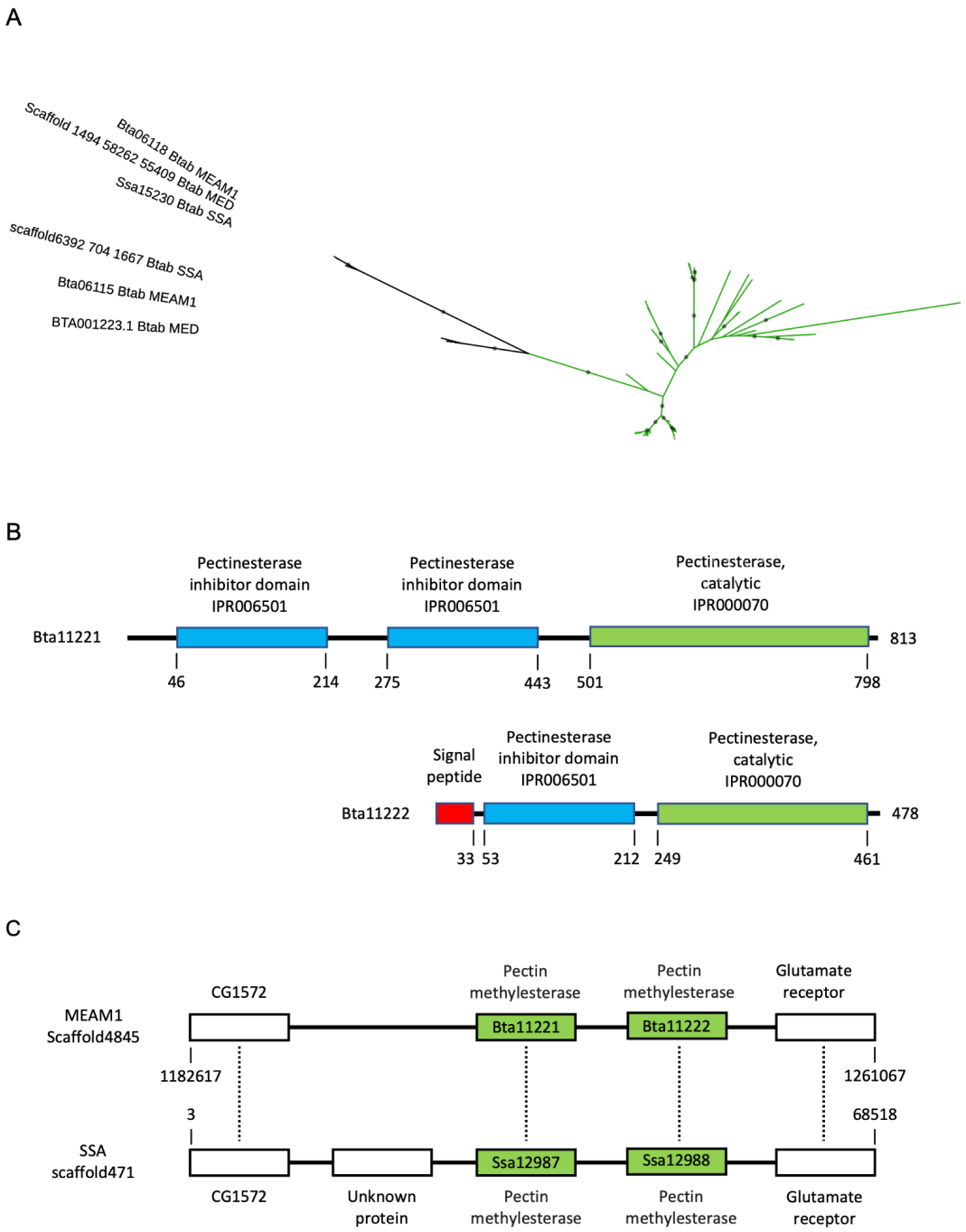
1081



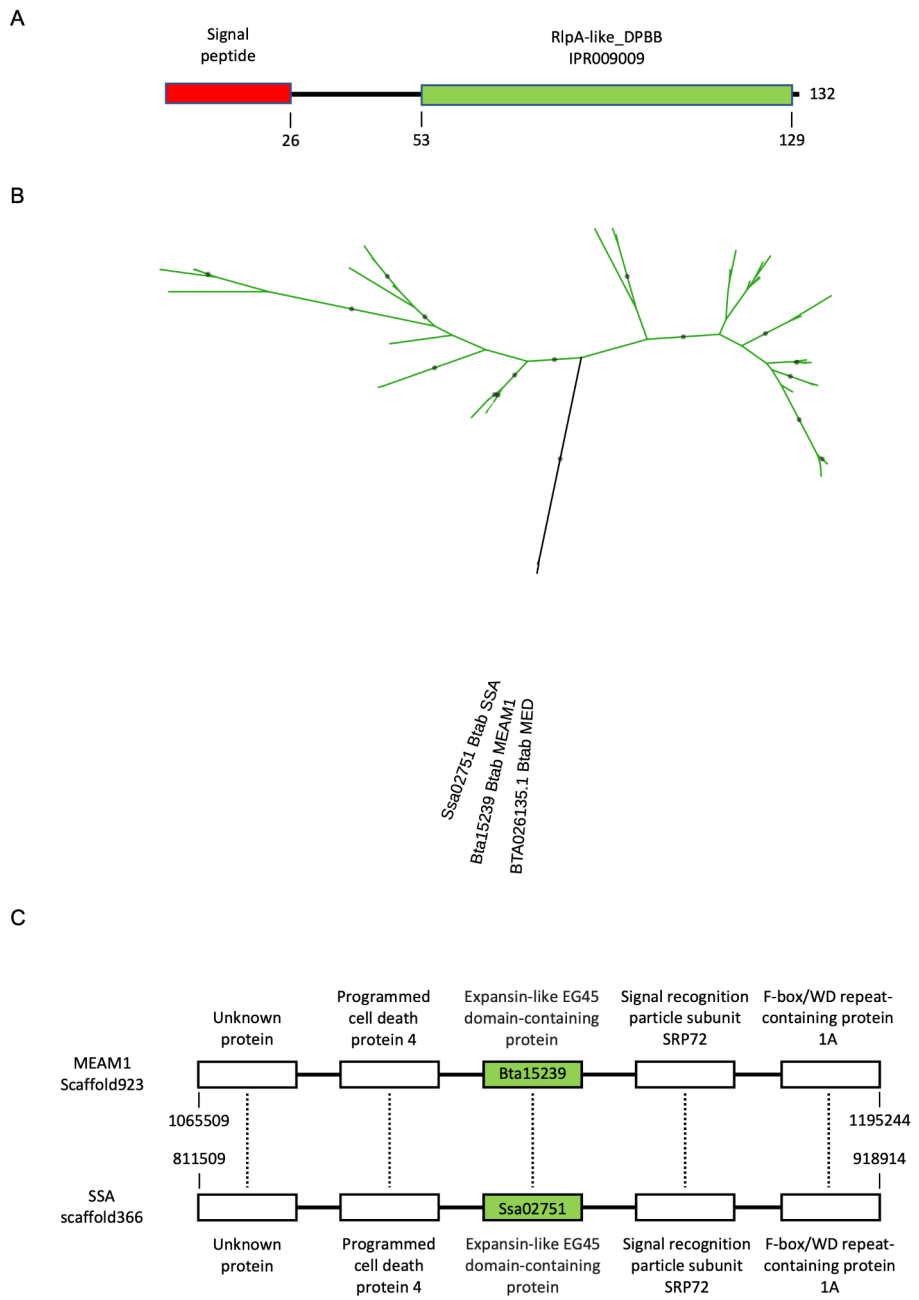
1082 **Figure 5.** Structural and functional analysis of the GH152 family protein Bta13961 acquired by HGT
1083 in *B. tabaci*. (A) Comparison of the crystal structure of tomato TL protein NP24-I (green) with the
1084 predicted structure of *B. tabaci* MEAM1 Bta13961 (cyan). Amino acids of the REDDD motif are
1085 colored red for the two putative catalytic residues and yellow for the other three. (B and C) HPAEC-
1086 PAD chromatograms of reaction products released by the GH152 Bta13961 from cellohexaose (G6)
1087 and laminarihexaose (L6) (B) and from barley glucan and lichenan (C).
1088
1089



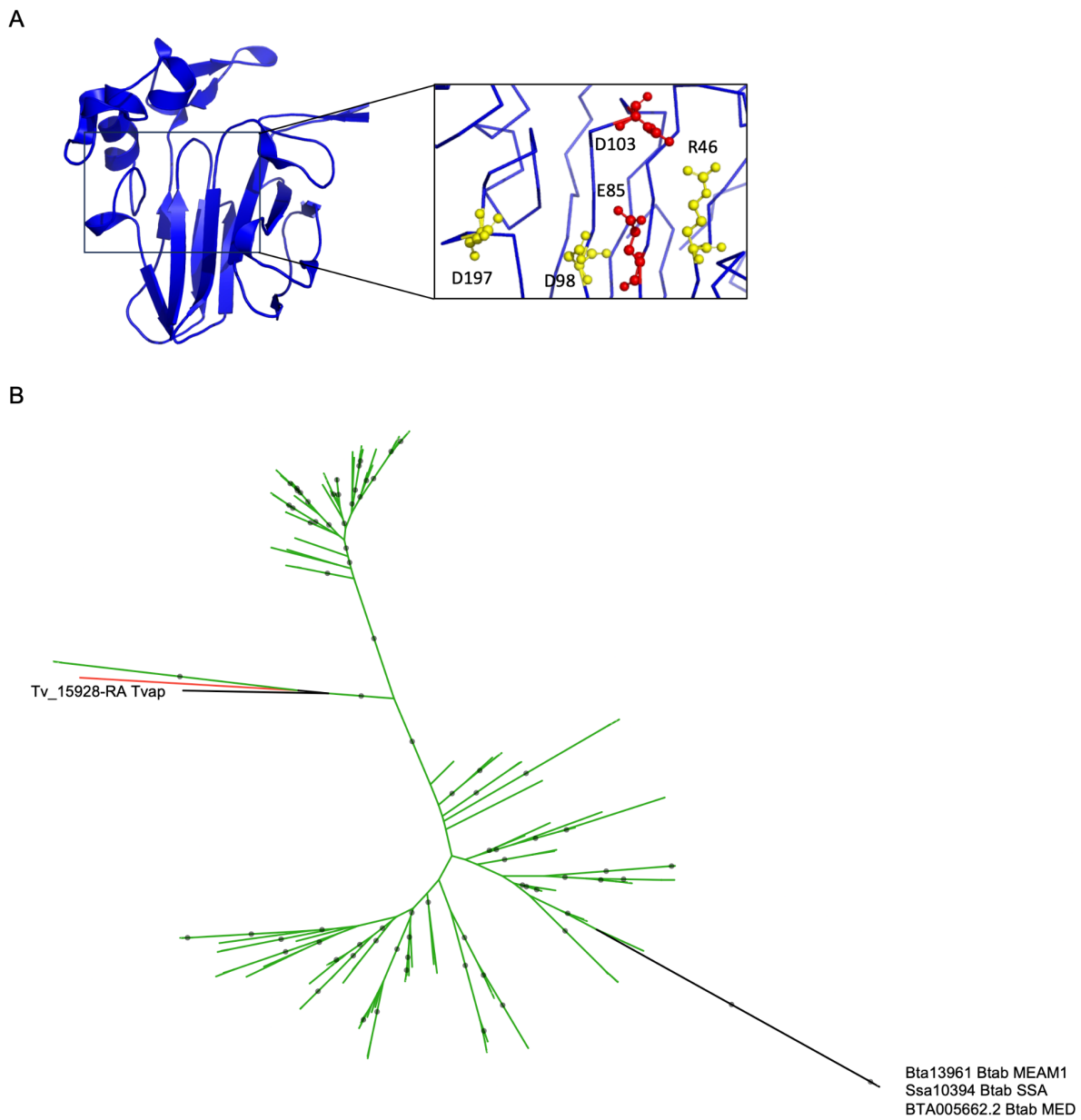
1090 **Figure 6.** Analysis of the pectin methylesterases acquired by HGT in *B. tabaci*. (A) Phylogenetic tree
1091 with *B. tabaci* MEAM1, MED and SSA sequences shown in black. Green branches represent plant
1092 sequences. Gray circles indicate nodes with support values greater than or equal to 80% and 95% for
1093 SH-aLRT and UFboot, respectively. (B) Domain organization of the *B. tabaci* MEAM1 Bta11221 and
1094 Bta11222 proteins. (C) Comparative map of the genomic region of the gene encoding the pectin
1095 methylesterases in *B. tabaci* MEAM1 (Bta11221 and Bta11222) and SSA (Ssa12987 and Ssa12988).
1096 Each rectangle represents a *B. tabaci* gene.



1097 **Figure 7.** Analysis of the expansin-related EG45 domain-containing protein acquired by HGT in *B.*
1098 *tabaci*. (A) Domain organization of the *B. tabaci* MEAM1 Bta15239 protein. (B) Phylogenetic tree with
1099 *B. tabaci* MEAM1, MED and SSA sequences shown in black. Green branches represent plant
1100 sequences. Gray circles indicate nodes with support values greater than or equal to 80% and 95% for
1101 SH-aLRT and UFboot, respectively. (C) Comparative map of the genomic region of the gene encoding
1102 the expansin-related EG45 domain-containing protein in *B. tabaci* MEAM1 (Bta15239) and SSA
1103 (Ssa02751). Each rectangle represents a *B. tabaci* gene.
1104



1105 **Figure 8.** Analysis of the GH152 family protein acquired by HGT in *B. tabaci* and *T. vaporariorum*.
1106 (A) Predicted structure of *T. vaporariorum* Tv_15928-RA colored in blue. Amino acids of the REDDD
1107 motif are colored red for the two putative catalytic residues and yellow for the other three. (B)
1108 Phylogenetic tree showing *B. tabaci* and *T. vaporariorum* sequences in black. Green branches represent
1109 plant sequences, and the red branch corresponds to a Metazoan sequence (*T. palmi* of the Thysanoptera
1110 order, accession number XP_034247376.1). Gray circles indicate nodes with support values greater
1111 than or equal to 80% and 95% for SH-aLRT and UFboot, respectively.
1112
1113
1114



1115 **Tables**

1116 **Table 1.** Overall comparison of HGT candidate search results

1117

	Bemisia tabaci MEAM1	*Trialeurodes vaporariorum*	*Frankliniella occidentalis*	*Thrips palmi*
Number of protein-coding genes	15,662	18,275	15,678	14,332
Number of predicted proteins with AHS > 0	**686**	**910**	**281**	**233**
Number of HGT candidates detected by AvP	**357**	**182**	**50**	**57**
- Bacteria	147	87	15	6
- Fungi	93	10	10	29
- Viridiplantae	78	15	7	9
- Other	39	70	18	13
Number of validated HGT candidates	**255**	**75**	**23**	**20**
- Bacteria	113	61	10	6
- Fungi	88	8	3	2
- Viridiplantae	53	5	1	3
- Complex: Bacteria or Fungi	1	1	10	9

1118

1119

1120

1121

1122 **Table 2.** Horizontally acquired CAZymes in *B. tabaci* and *T. vaporariorum.*

1123

CAZy		Origin of donor sequences	*B. tabaci*		*T. vaporariorum*	
Description	**Definition**		**HGT event**	**Sequence name**	**HGT event**	**Sequence name**
GH32	Glycoside Hydrolase Family 32 protein	Bacteria	BtaB53	Bta00773 Bta04549	-	-
CBM50	Carbohydrate-Binding Module Family 50	Bacteria	BtaB62	Bta15714	-	-
GH13	Glycoside Hydrolase Family 13 protein	Bacteria	-	-	TvaB10	Tv_04079-RA
GH49	Glycoside Hydrolase Family 49 protein	Fungi	BtaF06	Bta02250	-	-
GH71	Glycoside Hydrolase Family 71 protein	Fungi	BtaF15	Bta02866	-	-
CBM32-AA5_2	Carbohydrate-Binding Module Family 32 / Galactose oxidase	Fungi	BtaF18	Bta01072 Bta03567 Bta05090 Bta05091 Bta05092 Bta05093 Bta07691 Bta09066 Bta12453	-	-
GH30_3	Glycoside Hydrolase Family 30	Fungi	BtaF20	Bta02658	-	-
GH152	Glycoside Hydrolase Family 152 protein	Viridiplantae	BtaV01	Bta13961	TvaV01	Tv_15928-RA
GT10	Glycosyltransferase Family 10 protein	Viridiplantae	BtaV10	Bta08851	-	-
GH17	Glycoside Hydrolase Family 17 protein	Viridiplantae	BtaV12	Bta06115 Bta06118	-	-
EXPN	Distantly related to plant expansins	Viridiplantae	BtaV13	Bta15239	-	-
CE8	Carbohydrate Esterase Family 8 protein	Viridiplantae	BtaV24	Bta11221 Bta11222	-	-
GT61	Glycosyltransferase Family 61 protein	Viridiplantae	BtaV25	Bta14885	-	-
GT17	Glycosyltransferase Family 17 protein	Viridiplantae	BtaV27	Bta15659	-	-

PL1_4	Polysaccharide Lyase Family 1 protein	Complex: Bacteria or Fungi	BtaC01	Bta02634	TvaC01	Tv_05105-RA

1124

1125 **Table 3.** Horizontally acquired CAZymes in *F. occidentalis* and *T. palmi.*

1126

CAZy		Origin of donor sequences	*F. occidentalis*		*T. palmi*	
Description	**Definition**		**HGT event**	**Sequence accession**	**HGT event**	**Sequence accession**
GH32	Glycoside Hydrolase Family 32 protein	Bacteria	FocB01	XP_026273122 XP_026287851	TpaB01	XP_034233420 XP_034232383
GH5_8	Glycoside Hydrolase Family 5 protein	Bacteria	FocB02	XP_026276666 XP_026285289 XP_026285291	TpaB02	XP_034242052 XP_034242578
GH45	Glycoside Hydrolase Family 45 protein	Fungi	FocF01	XP_026287984 XP_026289264 XP_026291890	TpaF01	XP_034236588 XP_034233350
GH152	Glycoside Hydrolase Family 152 protein	Viridiplantae	FocV01	XP_026290383	TpaV01	XP_034247376 XP_034247673
CBM43-CBM43-CBM43-CBM43	Carbohydrate-Binding Module Family 43 protein	Viridiplantae	-	-	TpaV02	XP_034245819
PL1_4	Polysaccharide Lyase Family 1 protein	Complex: Bacteria or Fungi	FocC01	XP_026274115 XP_026279513 XP_026279528 XP_026280402 XP_026285066 XP_026285067 XP_026285883 XP_026289641 XP_026292396 XP_026292397	TpaC01	XP_034235640 XP_034237191 XP_034239440 XP_034241038 XP_034248354 XP_034248552 XP_034253769 XP_034253771 XP_034254938

1127

1128