


RESEARCH ARTICLE

Open Access

Lignocellulose degradation in isopods: new insights into the adaptation to terrestrial life



Marius Bredon, Benjamin Herran, Baptiste Lheraud, Joanne Bertaux, Pierre Grève, Bouziane Moumen and Didier Bouchon* 

Abstract

Background: Isopods constitute a particular group of crustaceans that has successfully colonized all environments including marine, freshwater and terrestrial habitats. Their ability to use various food sources, especially plant biomass, might be one of the reasons of their successful spread. All isopods, which feed on plants and their by-products, must be capable of lignocellulose degradation. This complex composite is the main component of plants and is therefore an important nutrient source for many living organisms. Its degradation requires a large repertoire of highly specialized Carbohydrate-Active enZymes (called CAZymes) which are produced by the organism itself and in some cases, by its associated microbiota. The acquisition of highly diversified CAZymes could have helped isopods to adapt to their diet and to their environment, especially during land colonization.

Results: To test this hypothesis, isopod host CAZomes (i.e. the entire CAZyme repertoire) were characterized in marine, freshwater and terrestrial species through a transcriptomic approach. Many CAZymes were identified in 64 isopod transcriptomes, comprising 27 de novo datasets. Our results show that marine, freshwater and terrestrial isopods exhibit different CAZomes, illustrating different strategies for lignocellulose degradation. The analysis of variations of the size of CAZy families shows these are expanded in terrestrial isopods while they are contracted in aquatic isopods; this pattern is probably resulting from the evolution of the host CAZomes during the terrestrial adaptation of isopods. We show that CAZyme gene duplications and horizontal transfers can be involved in adaptive divergence between isopod CAZomes.

Conclusions: Our characterization of the CAZomes in 64 isopods species provides new insights into the evolutionary processes that enabled isopods to conquer various environments, especially terrestrial ones.

Keywords: Transcriptomics, CAZymes, Isopods, Lignocellulose, Terrestrialization

Background

Plant biomass is the most abundant source of renewable carbon on earth [1]. Mostly composed of lignocellulose, it is an important resource for many organisms. Its decomposition involves the combined action of fungi, microbes and decomposer animals such as “litter transformer” macroarthropods [2, 3]. Among them, terrestrial isopods (Oniscidea) are known to contribute directly to litter decomposition and nutrient cycling by

digesting substrates [4–9], and indirectly through their faeces which affect the soil microbial community and its activity [10–12]. From marine ancestors, terrestrial isopods have successfully colonized all environments including freshwater and terrestrial habitats [13]. It is assumed that their ability to use plant biomass as a food source facilitated their colonization of terrestrial environments in the Late Paleozoic (~300 Ma, Permian–Carboniferous), together with morphological and physiological adaptations [14–17]. Their digestive tract consists of a short foregut comprising an esophagus and a stomach, a hindgut and a hepatopancreas where endogenous digestive enzymes are secreted; it allows an efficient

* Correspondence: didier.bouchon@univ-poitiers.fr; <http://ebi.labo.univ-poitiers.fr/>

Laboratoire Ecologie et Biologie des Interactions - UMR CNRS 7267, Equipe Ecologie Evolution Symbiose - Bâtiment B8-B35, Université de Poitiers, 5 rue Albert Turpain, TSA 51106, F-86073 Poitiers Cedex 9, France



digestion of their food [18, 19]. Lignocellulose constitutes the main food not only of terrestrial isopods, but also of freshwater isopods (*Asellota*) which feed on plant detritus of terrestrial origin [20, 21]. Many marine isopods (e.g. *Valvifera*, *Sphaeromatidea* and *Limnoriidae*) also consume cellulose and hemicellulose that arises, for example, from algae or driftwood [17].

Few studies have characterized enzymes which participate in lignocellulose decomposition in isopods. Kern et al. [22] have characterized an exoglucanase in the marine isopod *Limnoria quadripunctata* and Kostanjek et al. [23] have highlighted the endogenous production of an endoglucanase in the terrestrial isopod *Porcellio scaber*. These enzymes, identified as Carbohydrate Active enZymes (CAZymes) and classified as GH7 and GH9 in the families of the CAZy database [24], belong to a complex pathway involving many CAZymes that can degrade and release monosaccharides from lignocellulose. Recently, Bredon et al. [25] have identified 17 endogenous lignocellulose-degrading CAZy families in the common pill-bug *Armadillidium vulgare*, which illustrates the complexity of this process in this terrestrial isopod. Furthermore, some marine isopods could use hemocyanins in their digestive tract to modify lignin during lignocellulose digestion, thus facilitating access to cellulose and hemicellulose [26, 27].

However, there is so far no animal genome known to encode all necessary CAZymes to degrade lignocellulose, and in most cases, animals benefit from mutualistic associations with microbial symbionts allowing an efficient degradation of lignocellulose [28]. In arthropods like termites [29, 30] and beetles [31, 32], there is a complementary and synergistic action of the lignocellulose-degrading enzyme repertoire from the host and its associated microbial symbionts. The host and its microbiota achieve the degradation of the three lignocellulose components (i.e. cellulose, hemicellulose and lignin) in a cooperative manner in different parts of the digestive system. This is also the case for terrestrial isopods (*Oniscidea*), which interact with their microbiota to digest lignocellulose [16, 33–37]. This microbiota is mostly composed of hepatopancreas-resident bacteria and environmental bacteria localized in the hindgut [25]. It completes the set of lignocellulose-degrading enzymes that the host produces mostly in the hepatopancreas. Both marine and freshwater isopods belonging to the suborder *Asellota* share this ability with terrestrial isopods: they contain hepatopancreatic bacteria that contribute to the digestion of their food [20, 36]. However, several studies show that some marine isopods belonging to the suborders *Valvifera*, *Sphaeromatidea* and *Limnoriidea*, digest cellulose without the help of hepatopancreatic and gut bacteria [22, 36–39]. Strategies for lignocellulose digestion may therefore differ between

isopods species according to their biotopes and interactions between hosts and their microbiota. Moreover, the ability to digest lignocellulose varies in isopods, the terrestrial species being the most efficient ones [20]. Thus, the host CAZome (i.e. the CAZyme repertoire encoded in the host's genome) is expected to be different across isopods, according to their evolutionary history.

In the present study, we identified and annotated the host CAZomes of many isopod species belonging to different families to better understand their successful colonization of all environments. To this end, we identified CAZymes from 64 isopod transcriptomes, including both publicly available transcriptomes and 27 de novo sequenced and assembled datasets. These transcriptomes were obtained from a wide range of isopod species among *Oniscidea*, *Asellota*, *Valvifera*, *Sphaeromatidea* and *Limnoriidea*. Our analysis of the CAZomes from this large dataset highlights the different strategies for lignocellulose digestion in isopods, according to their adaptation to marine, freshwater and terrestrial environments. In addition, we assessed the molecular evolution of some lignocellulose-degrading key genes, suggesting an ancient origin and acquisition of these genes during the conquest of land.

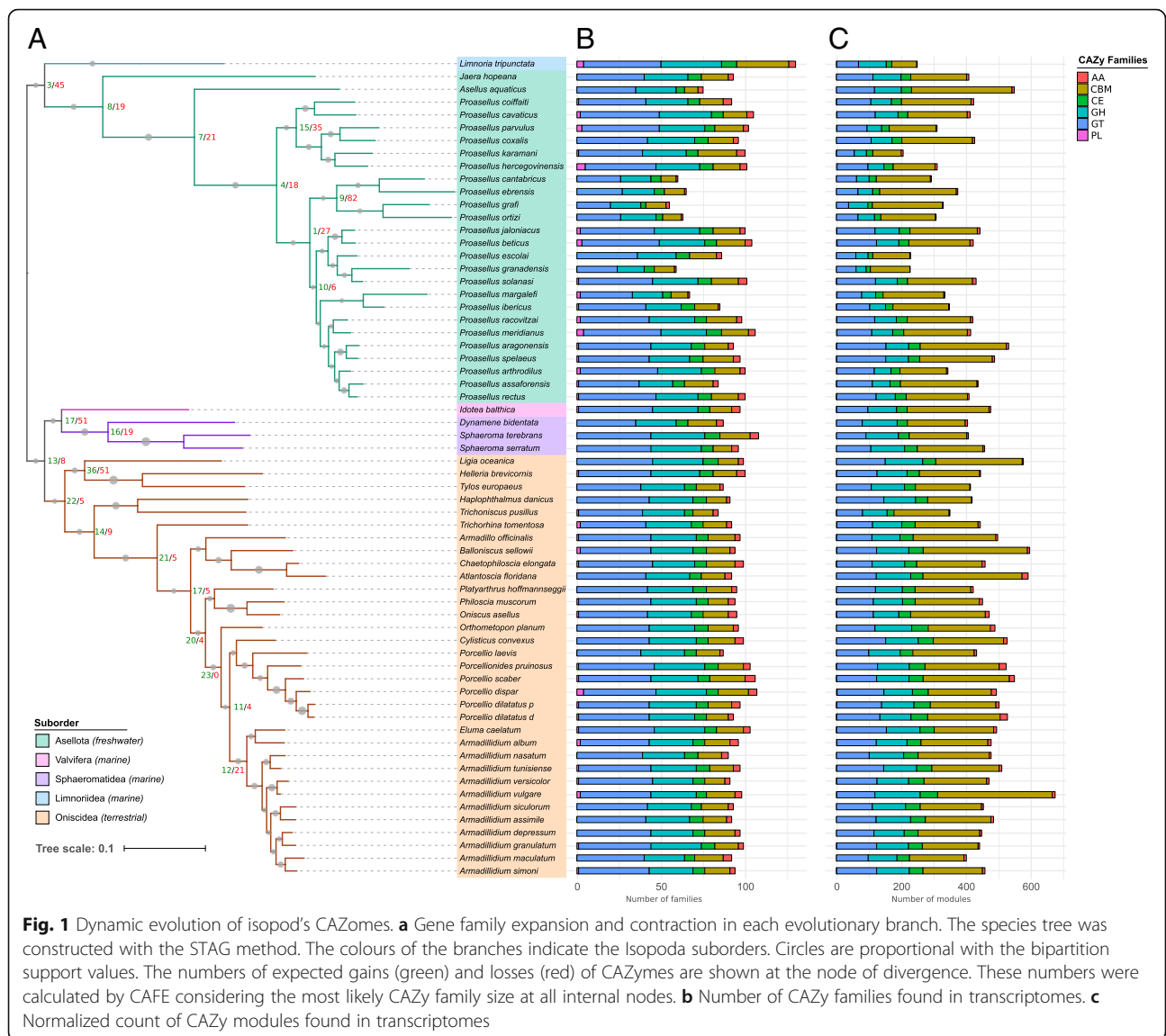
Results

Transcriptome assemblies

In total, 64 transcriptomes of isopod species, including 27 new transcriptomes, were assembled. The resulting assemblies ranged from 28,393 (*Jaera hopeana*) to 685,588 (*Limnoria tripunctata*) transcripts depending on the species (Additional file 1). Assembly completeness was evaluated using the BUSCO pipeline. Except for some *Proasellus* species, all the transcriptomes displayed a good completeness since more than 80% of the complete genes from the arthropod core genome were present in most of the assemblies (Additional file 2).

Distribution of the CAZy families

In total, 205 CAZy families were identified with an average of 93 families *per* isopod species (Additional files 3 and 4). Carbohydrate-Binding Modules (CBMs), Carbohydrate Esterases (CEs), Glycoside Hydrolases (GHs) and Glycosyl Transferases (GTs) were present in all isopod transcriptomes (Fig. 1). The CBM modules were the most numerous (334 modules on average per transcriptome), followed by the GT, GH and CE modules (on average 184, 135 and 59 per transcriptome, respectively). Polysaccharide Lyases (PLs) were absent in one third of the transcriptomes whereas Auxiliary Activities (AAs) were absent from the transcriptomes of *Proasellus margalefi*, *Proasellus grafi*, *Proasellus ebrensis* and *Proasellus cantabricus* (Fig. 1). When they were present, they were in small numbers (on average 8 and 1 modules per



transcriptome, respectively). Despite this feature which might be due to the low coverage in the transcriptomes involved, the distribution of the CAZymes appeared quite homogeneous across isopods. However, multivariate analysis (PCA) showed that CAZomes were different between isopod suborders (Fig. 2a) as only 84 CAZY families were shared by all isopod species. The transcriptomes of Asellota, Limnoriidea, Oniscidea, Sphaeromatidea and Valvifera contained a lot of specific CAZY families.

Asellota, Valvifera and Sphaeromatidea were dominated by CAZyme loss, whereas Oniscidea were dominated by CAZyme gains (Fig. 1, Additional file 4). The terrestrial isopods (Oniscidea) clade showed 22 expanded and 5 contracted CAZY families compared to the ancestral birth and death rate of CAZY families predicted for the node grouping Valvifera, Sphaeromatidea and Oniscidea. In comparison, the node grouping

Valvifera and Sphaeromatidea marine isopods showed 17 expanded and 51 contracted families. Similarly, the node grouping the marine isopod *Limnoria tripunctata* and Asellota freshwater isopods had larger numbers of contracted CAZY families than expanded ones (45 and 3 respectively) compared to the common ancestor of the 5 suborders. These trends were continued since Oniscidea showed a high number of expanded families compared to contracted families at almost all internal nodes, while Asellota were dominated by CAZY families' contractions. Interestingly, among the 31 rapidly evolving CAZY families predicted by CAFE (Family-wide P -value < 0.01; Additional file 4), seven are known to participate to the lignocellulose degradation, including hemicellulases (CE3, GH29, GH31, GH35, GH38), and some members of GH30s that are recognized as cellulases or hemicellulases, and cellobiose dehydrogenases (AA3).

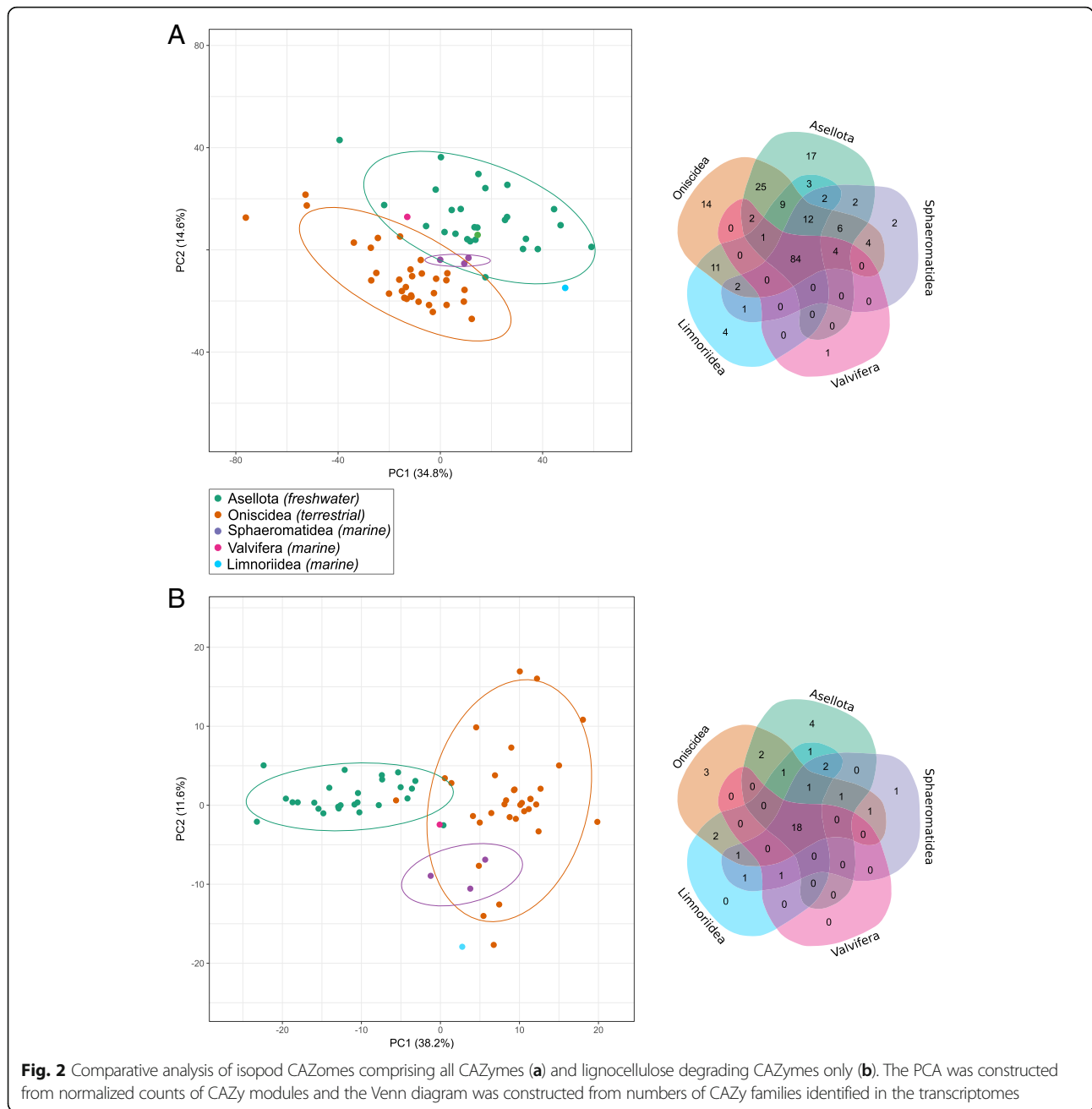
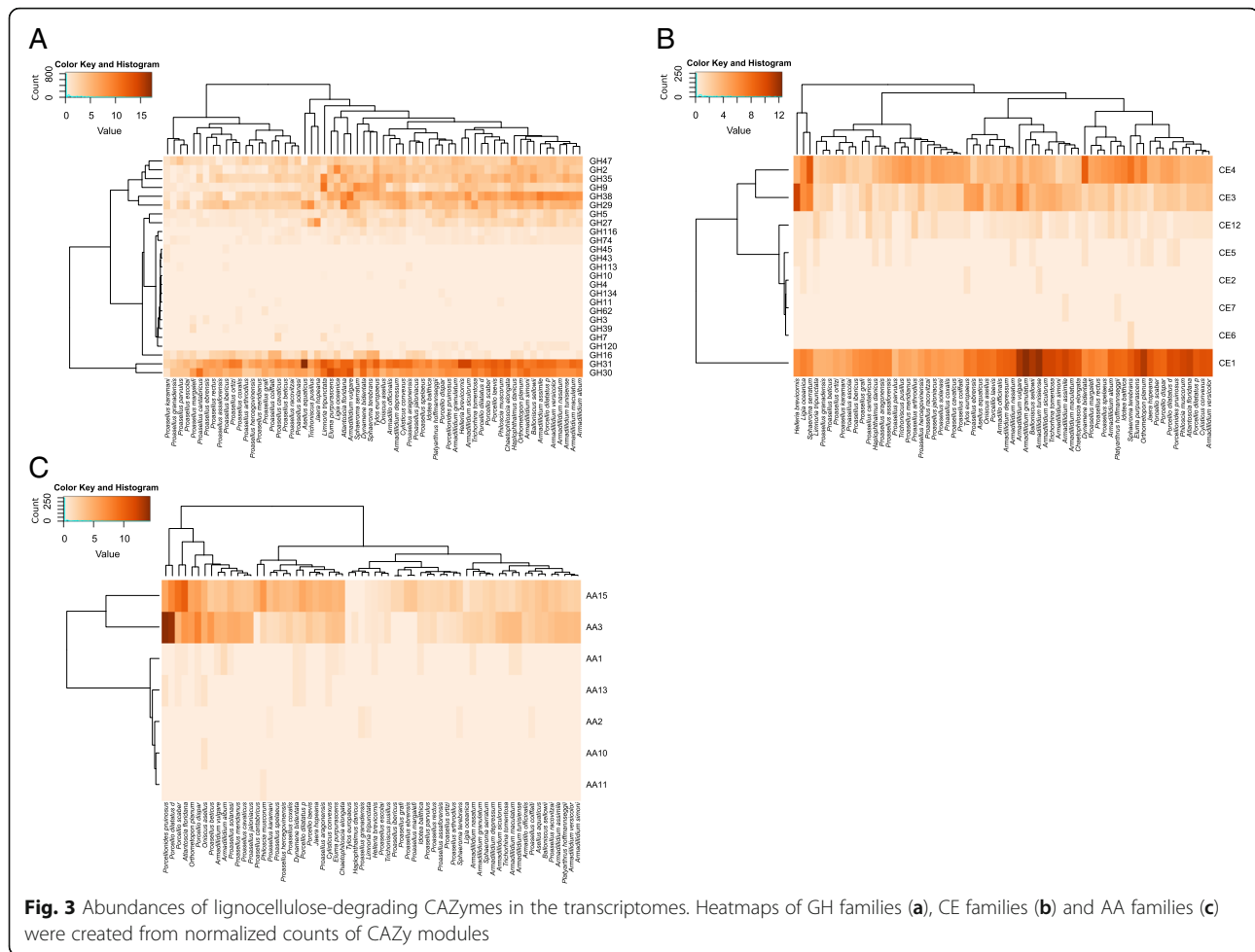


Fig. 2 Comparative analysis of isopod CAZomes comprising all CAZymes (a) and lignocellulose degrading CAZymes only (b). The PCA was constructed from normalized counts of CAZy modules and the Venn diagram was constructed from numbers of CAZy families identified in the transcriptomes

Lignocellulose degrading CAZymes in isopods

Given that not all CAZymes participate in lignocellulose degradation, lignocellulose degrading CAZymes were then examined in depth. Forty CAZy families were identified as lignocellulose degrading CAZymes in our transcriptomes, comprising AAs, GHs and CEs (Fig. 3). Among them, 18 were found in all isopod suborders. Despite this apparent functional redundancy, CAZomes linked to lignocellulose degradation were different across isopod species (Fig. 2b). Some CAZymes, including cellulases (GH5, GH9, GH30) and hemicellulases (GH27, GH29, GH30, GH31, GH35, GH47, CE1, CE3 and CE4), were found in more than 95%

of the transcriptomes. The families GH31, GH30 and CE1 were the most abundant ones (Fig. 3a-b, Additional file 4), with 1088, 905 and 823 modules respectively identified in the transcriptomes. Lignin modifying enzymes (LMEs) and oxidative cellulases belonging to AA families were rarer in isopods; the most abundant of these AA families was AA3 with 333 modules distributed in 57 species, followed by AA1 with 18 modules identified in 16 species (Fig. 3c, Additional file 4). Additionally, we identified 342 modules distributed among 61 species belonging to AA15, a CAZy family recently characterized in arthropods [40]. In many arthropods AA15s are likely involved in chitin modification,



but some of them could have expanded this family for cellulose digestion [40].

The prediction of enzyme activities with Hotpep showed that cellulases and hemicellulases are well distributed in isopod species (Fig. 4). Cellulases were found in almost all transcriptomes, notably endo- β -1,4-glucanases (EC 3.2.1.4) that were predicted in more than 93% of the transcriptomes and β -glucosidases (EC 3.2.1.21) in 76% of the transcriptomes. Similarly, hemicellulases, mannases (EC 3.2.1.22) and xyloglucanases (EC 3.2.1.23 and EC 3.2.1.51) were predicted in more than 95% of the transcriptomes. In contrast, LMEs and oxidative cellulases seemed to be exclusive to Oniscidea and Sphaeromatidea. Cellobiose dehydrogenases (EC 1.1.99.18) and laccases (EC 1.10.3.2) were predicted in most of terrestrial and marine isopods, and were not found in freshwater isopods, except for *P. jaloniacus* and *P. parvulus*.

GH9 and AA3 phylogenies

The GH9 enzymes are well studied in arthropods and many of them are known to participate in cellulose degradation. By contrast, AA3 enzymes are of interest

because they are not widespread in animals and their role in lignocellulose degradation remains vague in arthropods. The phylogenies of AA3 and GH9 genes show that the CAZymes from different isopod suborders were grouped as distinct clusters (Fig. 5). In addition, the sequences belonging to a given species were spread across several branches of the tree as discrete sub-clusters, suggesting different origins. The alignments of nucleotide sequences from a given species showed a percent identity lower than 80% on average. This divergence is too small to correspond to alternative splicing events and suggests instead multiple gene duplications that occurred repeatedly within the isopod order.

The AA3 genes showed significant (BUSTED likelihood ratio test, p -value < 0.05) signatures of gene-wide episodic diversifying selection in the whole phylogeny (252 sequences) as well as in the subtree (131 sequences) comprising all the AA3 genes from terrestrial isopods (Additional file 5). Furthermore, all sequences showed evidence of positive selection at individual sites. SLAC found evidence of pervasive positive selection on one site (#1128, p -value = 0.088, Additional file 5) and

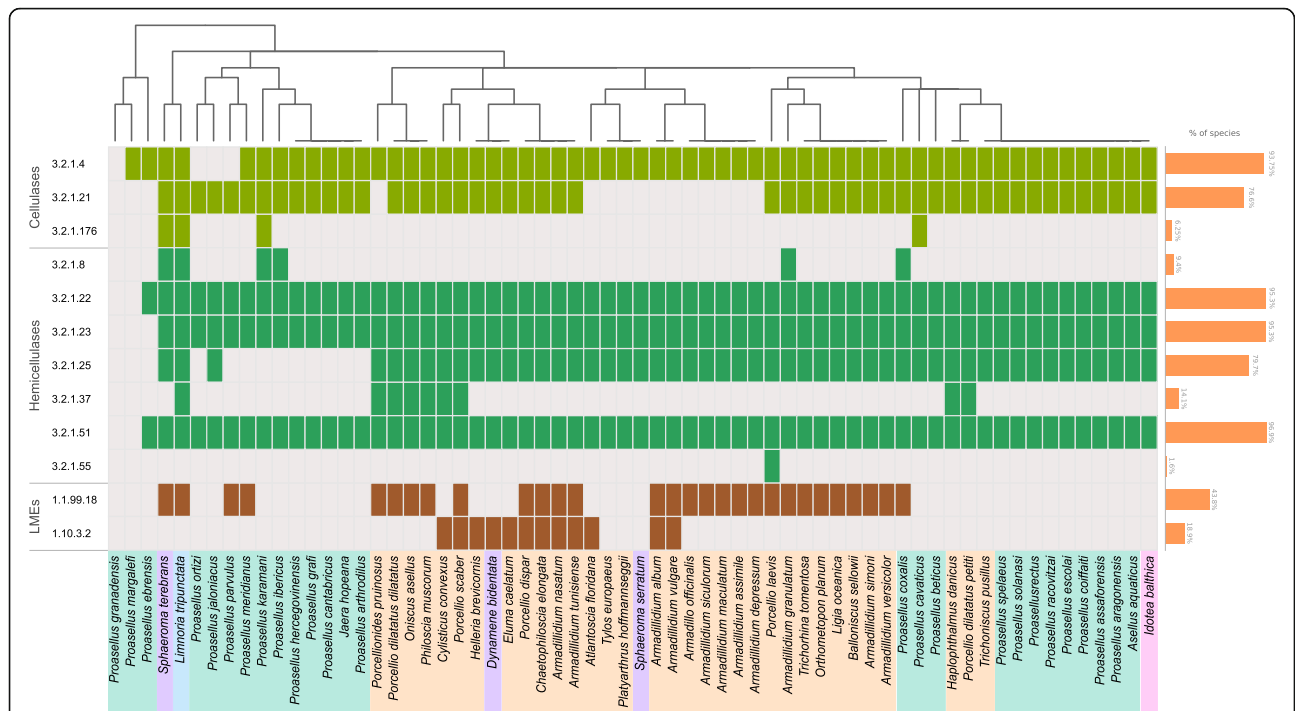


Fig. 4 Prediction of enzymatic functions (EC number) of cellulases, hemicellulases and LMEs identified in the transcriptomes. Coloured squares indicate the presence of a given enzymatic function in a transcriptome. Horizontal bars show the proportion of all transcriptomes with the presence of each enzymatic function

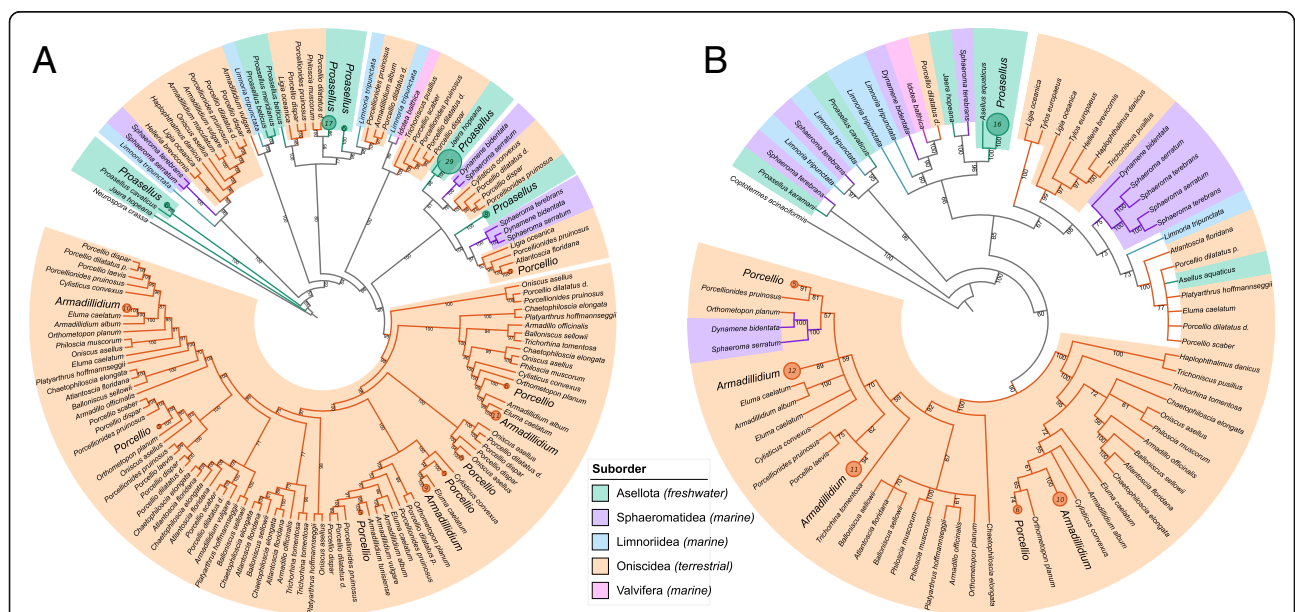


Fig. 5 AA3 (a) and GH9 (b) gene phylogenies. The colours of the branches indicate the Isopoda suborder. Fast bootstrap values ($n = 1000$) for branches are shown as percentages. To facilitate reading, circles materialize collapsed branches and the numbers within refer to the number of collapsed branches. The corresponding species were abbreviated through their genera (*Proasellus* sp.; *Porcellio* sp.; *Armadillidium* sp.)

negative selection on 630 sites (Additional file 5) whereas FUBAR found positive selection on 5 sites (#957, #1022, #1042, #1059 and #1076, posterior probability > 0.9, Additional file 5) and negative selection on 649 sites among 1277.

Unlike AA3 there is no evidence that any sites have experienced diversifying selection in the branches of the GH9 phylogeny (Additional file 5). We did not identify any site under positive selection using SLAC method with a *p*-value threshold of 0.1 (Additional file 5) whereas 139 sites among 178 were under negative selection. FUBAR found evidence of positive selection on one site (#12, posterior probability = 0.992, Additional file 5) and negative selection on 164 sites among 178.

GH7 phylogeny

Protein BLAST searches against the non-redundant protein database of NCBI using GH7 proteins identified in isopod transcriptomes showed a large number of matches with GH7 sequences isolated from fungi (basidiomycetes and ascomycetes), and apart from crustaceans,

there was no match with GH7 from any other animals. The comparison of isopod GH7 protein sequences with their homologues in other crustaceans, fungi, oomycetes, protists, amoebzoa and demosponges, showed that they were nested in a clade comprising all the proteins of crustaceans and oomycetes (Chromista), and some proteins of ascomycetes (Fig. 6). This clade was separated from the other clades formed by the GH7 from protists, ascomycetes and basidiomycetes.

Discussion

The acquisition of lignocellulose degrading CAZymes in isopods could have helped them to better use their diet and favoured their adaptation to the environment, especially during land colonization. The 64 isopod transcriptomes analysed in this study enabled the identification of 205 CAZy families in total and 40 lignocellulose degrading CAZymes, including endogenous cellulases, hemicellulases and LMEs. The resulting CAZomes identified from various freshwater, marine and terrestrial isopod species, highlight different repertoires. Isopods have a

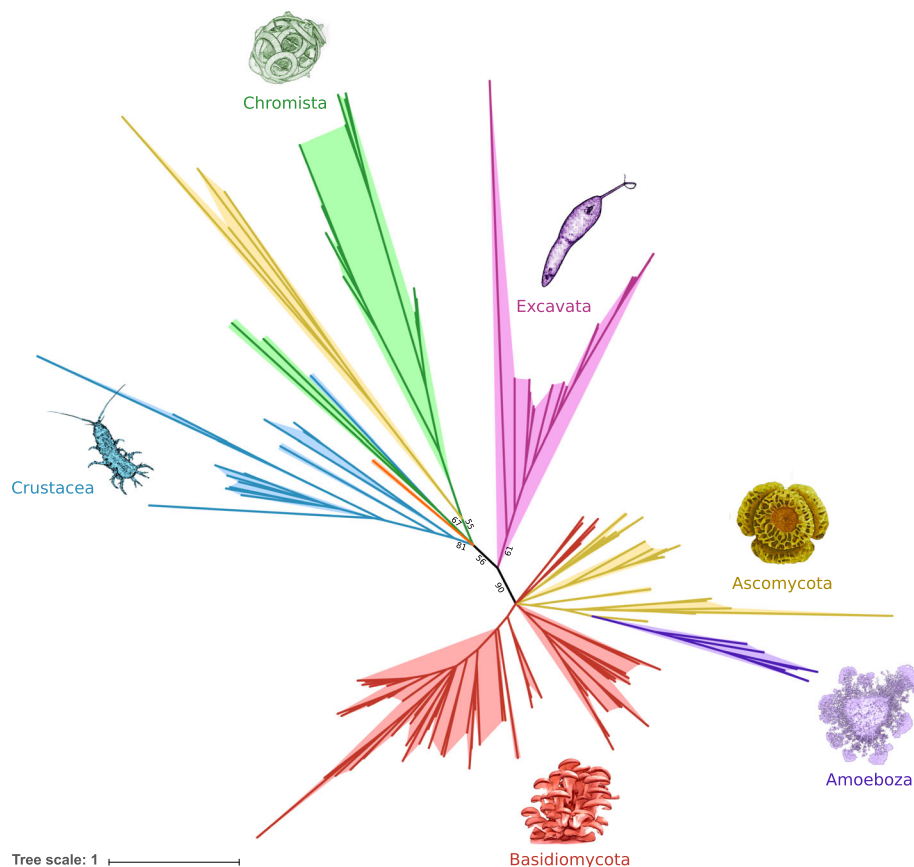


Fig. 6 Phylogenetic analysis of the GH7 family from isopods. An unrooted tree shows the relationships between the GH7 proteins of Crustacea (in blue), Demospongea (in orange), Amoebzoa (in purple), Basidiomycota (in red), Excavata (in pink), Chromista (in green) and Ascomycota (in yellow). All branches are drawn to scale as indicated by the scale bar. Fast bootstrap values ($n = 1000$) for the main branches are shown as percentages

great number of CAZymes, placing them at the same level as notorious decomposers such as termite or fungi in terms of CAZymes diversity [41, 42].

Previous studies have shown that the higher abundance of CAZymes in saprophytic fungi is linked to their plant-based nutrition, while lignocellulose degrading CAZymes are less abundant in endophytic fungi or fungi parasitizing animals [43, 44]. We showed that isopods have also been subjected to CAZyme gains and losses, especially as concerns lignocellulose degrading CAZymes. The number of gain events exceeded the number of loss events in terrestrial isopods, indicating that lignocellulose degrading CAZymes acquisition is dominant in this group. On the contrary, lignocellulose degrading CAZymes loss is dominant in aquatic isopods (suborders Asellota, Limnoriidea, Sphaeromatidea and Valvifera). Our results as well as previous studies strongly suggest that some endogenous cell-wall degrading enzymes have been acquired for a long time [40, 45, 46], explaining why some marine isopods could efficiently degrade cellulose and hemicellulose, and in most cases without the contribution of microbiota [26, 27]. Fossil records and phylogenomic analysis dated the origin of terrestrial isopods (suborder Oniscidea) at 300 Mya, which coincides with the diversification of vascular plants on land [47]. To deal with terrestrial plant cell walls that are relatively indigestible [48], a large enzymatic repertoire is necessary. Such conditions would have likely favored the expansion of endogenous CAZymes and the acquisition of digestion-enhancing microbiota in terrestrial isopods to cope with food resources diversification [16, 25, 49]. Gene acquisition seems to be an ancient but still ongoing process in the terrestrial isopods, given the high number of events we observed and the fact that most of them were distributed across the subclade grouping all Oniscidea in the phylogenetic tree. This suggests that a more complete lignocellulose degrading CAZymes repertoire was selected in the terrestrial species. In comparison, freshwater isopods have a less expanded CAZomes. However they could compensate by feeding on detritus that have been partially digested by microbial enzymes [20].

Cellulose and hemicellulose being common resources for most isopods, it is not surprising that both cellulases and hemicellulases were identified in all the transcriptomes. Only three hemicellulases (two xyloglucanases and one mannanase) were found in almost all transcriptomes. However, the degradation of hemicellulose, which has a variable composition both within and between plant tissues and species, requires a large enzymatic arsenal. Hence, the other enzymes necessary for this process in isopods could be brought by the microbiota, as shown in *A. vulgare* in which the microbiota plays the major role in hemicellulose degradation [25]. The

cellulases belonging to endo- β -1,4-glucanases (EC 3.2.1.4), found to be widespread in isopods, are classified in the GH9 family, an ancient and common eukaryotic gene family [45, 46]. We showed in this study that GH9 sequences are highly conserved in isopods. They were thus duplicated many times in a conservative manner in isopods allowing them to increase the number of expressed enzymes and thus to increase the rate of cellulose digestion. Several other crustacean species are known to express many GH9 genes, like the crayfish *Cherax quadricarinatus* [50] or the land crab *Gecarcoida natalis* [51] which possess twenty and three forms of GH9 respectively. Furthermore, we identified β -glucosidases (EC 3.2.1.21) affiliated to GH5 family in 49 out of 64 isopod transcriptomes, and lytic polysaccharide monoxygenases (AA15) in 61 transcriptomes. Lytic polysaccharide monoxygenases belong to a CAZY family known for its role in chitin remodeling, and they were recently demonstrated to be involved in cellulose oxidation in some arthropods [40]. This is the first record of lytic polysaccharide monoxygenases in isopods, and because several AA15 modules were identified in isopod transcriptomes, we might suppose that they were expanded in isopods to play a role in cellulose degradation like in some other arthropods. Thus, after a mechanical fragmentation of the food that facilitates enzyme access to lignocellulose [25], the synergistic action of high numbers of endo- β -1,4-glucanases, β -glucosidases and lytic polysaccharide monoxygenases would enable isopods to degrade and ingest the cellulose.

In addition to endoglucanases and glucosidases, exoglucanases, belonging to the GH7 family, were found in four isopod species: *Limnoria tripunctata*, *Proasellus cavaticus*, *Proasellus karamani*, and *Sphaeroma terebrans*. Exoglucanases are required by fungi to degrade the cellulose [52], but they are uncommon in other eukaryotes and prokaryotes. It is known that a small number of crustaceans [22, 26, 27, 53, 54] possesses exoglucanases, constituting rare examples in animals [24]. In our phylogenetic analysis, all the GH7s from crustaceans clustered with oomycete's GH7s, suggesting an ancient acquisition of exoglucanases in crustaceans, perhaps through horizontal gene transfers with oomycetes as putative donors [26]. Indeed, such transfers of lignocellulose degrading enzymes have already been shown in nematodes [55–57] and insects [58–61]. Alternatively, if the GH7 family was present in the common ancestor of all crustaceans, it was then lost in most isopods.

Lignin being restricted to terrestrial plants, terrestrial isopods are expected to have the capacity to break down this cross-linked phenolic polymer [16]. Its oxidation needs LMEs that are classified in AA families. These enzymes are widespread in fungi and bacteria [62], but very few animals are known to express LMEs. Terrestrial

isopods are able to undertake an oxidative degradation of the phenolic compounds [63, 64], and the expression of endogenous laccases (EC 1.10.3.2) and cellobiose dehydrogenases (CDH; EC 1.1.99.18) has been demonstrated in the hindgut of the pill bug *A. vulgare* [25]. Our results showed that CDH are widespread in all Oniscidea and rapidly expanding in this group, while laccases are restricted to some species. Cellobiose dehydrogenases belonging to the AA3 family, are lignin-degrading auxiliary enzymes that assist the activity of the other AAs or support the action of the GHs [62, 65]. They are unable to degrade lignin on their own, but they are known to be involved in the breakdown of cellulose in some fungi. On the other hand, CDH are not widespread in arthropods. They have not been identified in insects, where CAZymes belonging to AA3 family are actually implicated in growth and immunity [66, 67]. In fungi, the multiplicity of enzymatic functions of the AA3 family is reflected by the multigenicity of the AA3 genes [61]. Our study suggests that evolution of the AA3 genes is under episodic diversifying selection in isopods. Moreover, we showed the expansion of the AA3 family genes in Oniscidea: thus, gene duplication events followed by gene diversification might lead to the acquisition of CDH. Their role remains unclear, but they are very likely to participate to the lignocellulose degradation in terrestrial isopods.

In contrast, we showed that AAs are rarer in aquatic isopods and no LMEs were identified in most of the sampled transcriptomes. However, members of the Asellota should also have the capacity to oxidize lignin because they exploit the same food sources as most Oniscidea. The LMEs in Asellota could therefore be of microbial origin, as suggested by Zimmer and Bartholmé [20] upon showing that the activity of hepatopancreatic phenol oxidases is reduced after treatment with antibiotics. Most marine isopods feed on seaweeds, essentially composed of cellulose and hemicellulose but also of phenols (e.g. brown algae) for some of them [49, 68]. However the ability to oxidize phenols is not widespread in marine isopods; Zimmer et al. [17] showed that *Gnori-sphaeroma oregonense* (Sphaeromatidea) has the ability to oxidize phenols while *Idotea wosnesenskii* (Valvifera) has not, despite a diet rich in phenols. *Sphaeroma terebrans* is the only marine species where we predicted an endogenous CDH. While our data suggest that it could degrade the lignocellulose from the wood, Si et al. [69] showed that wood is an unlikely food source for this marine wood-boring isopod and serve only as a source of shelter. Its nutrition could derive from a microphagous filter-feeding habit. Interestingly, it was shown that marine wood-feeding isopods of the genus *Limnoria* (Limnoriidea) can degrade lignin with the help of its hemocyanins that are activated into

phenoloxidases [26, 27]. The presence of hemocyanin genes was investigated using sequence similarity searching, and unsurprisingly all transcriptomes show high numbers of hemocyanins. Nevertheless, our data do not allow us to conclude about their potential role in lignocellulose degradation. Hence, marine isopods seem to have a number of strategies to degrade phenols. The ability to degrade lignin could be an important pre-adaptation to terrestrial life [17, 49], which would explain why the ability to degrade phenols alone may have evolved independently in marine isopods.

Conclusion

Several strategies for lignocellulose degradation have evolved within isopod species. Marine, freshwater and terrestrial isopods showed different CAZomes, the latter having probably significantly evolved during the lifestyle transition from water to land. Similarly, isopod ecology and evolution could have driven the distribution of their endogenous CAZymes, and the microbiota associated to lignocellulose degradation could also influence and be influenced by host CAZomes. For example, a functional complementarity between the host and its microbiome for lignocellulose degradation has been highlighted in the terrestrial isopod *A. vulgare* [25]. Taking the hologenome concept into consideration (i.e. the sum of the host genome and its microbiome), variations in the hologenome can be due to changes in either the host genome or the microbiome [70]. From our results, it appeared clearly that terrestrial and freshwater isopods cannot digest lignocellulose by themselves. To respond to environmental changes, the microbiome can adjust more rapidly than its host, bringing a higher contribution to collective functions in the hologenome. It is therefore important to consider both the microbiome and the evolution of host CAZomes to understand the successful colonization of land by isopods.

Methods

Biological samples

Transcriptomic data of 48 isopod species were retrieved from SRA archive (Additional file 1). For each of them, we ascertained that all tissues were represented or that data were generated from whole individuals. Additionally, we generated transcriptomes from 27 isopod species sampled from our laboratory rearing or collected from various field sites in 2015 and 2016 (Additional file 1). By combining these newly generated data to already publicly available data, we finally obtained transcriptomes from 64 different isopod species including new transcriptomic data for 16 species.

RNA extraction and sequencing

To extract total RNA from the 27 collected isopod species, whole individuals were frozen in liquid nitrogen and ground with a mortar and pestle. Total RNA was then extracted using the RNeasy kit (Qiagen) and treated with RNase-free DNase I (Qiagen), according to the manufacturer's protocols. After quantification with NanoDrop™ technology, the RNAs were stored at -80 °C.

The 125 bp paired end sequencing of the extracted RNAs from 27 isopod species were performed on a HiSeq 2500 using Illumina technology (Additional file 1). Each library was constructed with the total RNA of a pool of five males and five females (except for *Trichoniscus pusillus* for which only one male could be sampled). The poly-A selection of the mRNA and the sequencing were carried out by Eurofins (<https://www.eurofinsgenomics.eu/>).

Transcriptome assembly

Read quality was checked with FastQC (version 0.11.2; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Removal of sequencing adaptors and low quality bases was performed with Trimmomatic (version 0.32; [71]). Reads shorter than 35 bp were discarded. Pre-processed reads from each species were assembled using IDBA-Tran [72] with default parameters. Transcript redundancy was removed by clustering with ≥95% identity using CD-HIT-EST (version 4.6; [73]). The completeness of the resulting assemblies was assessed with BUSCO (version 3.0.1; [74]) referring to core arthropod genes.

Carbohydrate-active enZYme annotation

CAZymes were identified using the Carbohydrate Active enZymes (CAZy) database [24]. Prior to identification, all open reading frames (ORFs) were predicted from transcriptomes using Transdecoder (version 3.0.1; <https://transdecoder.github.io/>) with default parameters. Subsequently, dbCAN [75], a database which uses hidden Markov Models to define the signature domains for each CAZy family (i.e. Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), Auxiliary Activities (AAs) and Carbohydrate-Binding Modules (CBMs)), was used to identify CAZymes. All predicted ORFs were analysed with dbCAN (October 1, 2017) using HMMER (version 3.1b2; [76]) with an E-value threshold of 0.0001.

To remove CAZymes originating from host microbiota including fungi, prokaryotes and viruses, ORFs identified as CAZymes were compared with the Non-Redundant Protein database (October 1, 2017) using BLASTP [77] with an E-value cut-off of 0.0001. The BLAST outputs were then imported into MEGAN6 software (version 6.9; [78]) for taxonomic assignment using the NCBI

taxonomy database. All ORFs assigned to fungi, prokaryotes or viruses were discarded. To verify that all non-host CAZymes were discarded, the 37,481 remaining ORFs were compared again with the latest Non-Redundant Protein database (April 25, 2019), using DIAMOND [79] with the "--more-sensitive" mode and an E-value cut-off of 0.001. Among them, 35,194 were assigned to metazoans by MEGAN6 software (version 6.9; [78]), 120 to eukaryotes, 488 to cellular organisms, and 8 were not taxonomically assigned. The 1671 remaining ORFs, most of them being CBMs (1515 including 1068 CBM14), have no hits in the Non-Redundant Protein database. We are aware that this method could exclude sequences that may have been the result of recent horizontal gene transfer. However, without any further genomic information we cannot identify such transfers. The resulting host CAZymes were then imported into Hotpep [80] to predict their enzymatic activity.

CAZyme counts (note that "CAZyme" refers to functional modules or domains, not genes) were normalized to compare their diversity across isopod species. This was done to even out the heterogeneity arising from differential transcriptomic sampling and sequencing methods. CAZyme counts for each transcriptome were divided by the number of ORFs in the transcriptome of interest to calculate the relative abundance for each CAZyme family. Then, normalized count of each family in each transcriptome was calculated by multiplying the relative abundance by 15,256, representing the lowest number of ORFs identified in our dataset corresponding to the *Proasellus ebrensis* transcriptome. In summary, the normalization was calculated as follows:

$$\frac{\text{number of CAZymes in a transcriptome of interest} \times \text{number of ORFs in the transcriptome of } P.\text{ebrensis}}{\text{number of ORFs in a transcriptome of interest}}$$

Normalized counts of CAZymes were then used for hierarchical clustering using R (version 3.4.0; <http://www.R-project.org/>) and Principal Component Analysis (PCA) using ClustVis [81].

The species tree was inferred with Orthofinder2 (version 2.2.7; [82]) according to the STAG (Species Tree Inference from All Genes) method that infers a species tree from sets of multi-copy gene tree [83]. Next, the variation (expansions and contractions) of the CAZyme gene family sizes were analysed using CAFE (version 4.2; [84]). CAFE estimates the global birth and death rate of CAZy families considering their sizes in the extant species and then infers the most likely gene family size at all internal nodes of the phylogeny. Only CAZy families that were shared by more than two species were kept for this analysis.

GH9 and AA3 phylogenies

Selected sequences which were present in most sampled isopods were examined in depth to infer their phylogeny. Enzymes belonging to GH9 and AA3 families were identified in 97 and 89% of the transcriptomes. All GH9 sequences used for the phylogeny were predicted by Hotpep to function as endo- β -1,4-glucanases (EC 3.2.1.4). Not all the activities of AA3 sequences were predicted and only 54 sequences were predicted to be cellobiose dehydrogenases (EC 1.1.99.18). Consequently, the AA3 sequences were compared to the Pfam database (version 32.0; [85]) using hmm-search (version 3.1b2; [76]) with an E-value cut-off of 0.00001 to identify conserved domains. In a conservative approach, all sequences lacking the conserved domain GMC oxidoreductase (an enzyme family that includes cellobiose dehydrogenases) were discarded. All remaining gene sequences (251 AA3 and 128 GH9) were aligned with Muscle [86] and the subsets of sites used for the phylogenetic analyses were determined with Gblocks using Less Stringent Selection parameters [87]. Phylogenies were constructed with the iqTree software (version 1.6.2; [88]) using ModelFinder for model selection [89] and performing 1000 ultrafast bootstrap with UFBoot [90]. For the construction of the phylogenetic trees, ModelFinder selected the GTR + F + R7 and the TIM2 + F + R6 substitution models for AA3 and GH9, respectively. Trees were modified with the iTol online tool [91]; branches were collapsed when the ultrafast bootstrap values were less than 50. When very close sequences of a given isopod species branched together in the tree, only one sequence was kept. The GH9 tree was rooted with an endo- β -1,4-glucanase from the termite *Coptotermes acinaciformis* (GenBank accession number: AAK12339.1) and the AA3 tree was rooted with a cellobiose dehydrogenase from the fungus *Neurospora crassa* (GenBank accession number: EAA28998.1).

Assessing positive selection

Multiple tests for selection were performed using the Datamonkey 2.0 web application [92] through i) gene-wide, ii) polymorphic-based and iii) codon-based approaches. The BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) test [93] was used to identify which branches of the phylogeny were under positive selection. We also performed a set of codon-based tests which aim to identify sites under selection pressure by estimating the rates of nonsynonymous (dN) and synonymous (dS) changes at each site in the sequence alignment: i) SLAC (Single-Likelihood Ancestor Counting method, [94]) that uses a combination of maximum-likelihood (ML) and counting approaches and ii) FUBAR (Fast Unconstrained Bayesian Approximation for inferring selection; [95]) that uses a Bayesian

approach, assuming that the selection pressure for each site is constant along the entire phylogeny.

GH7 phylogeny

Unlike GH9 and AA3, GH7 enzymes were identified in only four transcriptomes (*L. tripunctata*, *P. cavaticus*, *P. karamani*, and *S. terebrans*). To infer their putative origin, a phylogenetic tree was constructed with GH7 sequences from other organisms. A BLASTP using GH7 proteins identified in isopod transcriptomes as query was applied to identify GH7 homologues in the NR protein database of NCBI. The GH7 protein sequences were also retrieved from GenBank with GH7 Pfam id (pfam00840 and cl21662) as query and from the CAZY database (<http://www.cazy.org/>). Methods used for alignment and phylogenetic tree reconstruction were the same as described above. The WAG+F + R6 substitution model was selected for phylogenetic tree constructions. Branches were collapsed when the ultrafast bootstrap values were less than 50.

Additional files

Additional file 1: Metrics of samples and associated transcriptomes. (XLSX 19 kb)

Additional file 2: Assembly completeness assessed with BUSCO referring to core arthropod genes. (PNG 411 kb)

Additional file 3: Protein sequences of CAZY genes identified in the transcriptomes. (FA 16407 kb)

Additional file 4: List of CAZymes identified in the transcriptomes. (XLSX 64 kb)

Additional file 5: Gene-wide and individual site tests for positive selection assessed with BUSTED, SLAC and FUBAR. (XLSX 323 kb)

Abbreviations

AA: Auxiliary Activities; CAZyme: Carbohydrate-Active enzyme; CBM: Carbohydrate Binding Module; CDH: Cellobiose DeHydrogenase; CE: Carbohydrate Esterase; EC: Enzyme Commission number; GH: Glycoside Hydrolase; GT: Glycosyl Transferase; LME: Lignin modifying enzyme; ML: Maximum likelihood; NCBI: National Center for Biotechnology Information; ORF: Open reading frame; PCA: Principal Components Analysis; PL: Polysaccharide Lyase; RNA: RiboNucleic Acid; SRA: Sequence Read Archive

Acknowledgements

We thank Carine Delaunay for assistance in RNA extractions and Alexandra Lafitte for animal rearing.

Authors' contributions

DB and MB conceived and designed the study. DB and MB supervised the study. MB performed sequence processing, data analysis and drafted the manuscript. BH prepared RNA for sequencing and performed assemblies of the transcriptomes. JB and PG supervised the acquisition of transcriptomic data. BL participated to phylogenetic reconstructions. BM supervised the bioinformatic data analysis. All authors contributed to the final version of the manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by the 2015–2020 State-Region Planning Contracts (CPER), European Regional Development Fund (FEDER) (BiodivUP project, coordinator DB), and intramural funds from the Centre National de la Recherche Scientifique and the University of Poitiers. The funding agencies were not involved in designing the study nor in collecting, analysing, and

interpreting the data or in writing the manuscript. MB was supported by a PhD grant from the French Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.

Availability of data and materials

Reads used for transcriptomes assemblies are available from the NCBI Sequence Read Archive under accession numbers provided in the Additional file 1. Identified CAZymes are provided in FASTA format in the Additional file 2.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 11 March 2019 Accepted: 23 May 2019

Published online: 07 June 2019

References

- Swift M, Heal O, Anderson J. Decomposition in terrestrial ecosystems. Oxford: Blackwell Scientific Publications; 1979.
- David JF. The role of litter-feeding macroarthropods in decomposition processes: a reappraisal of common views. *Soil Biol Biochem*. 2014;76:109–18.
- López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P. Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. *Sci Rep*. 2016;6. <https://doi.org/10.1038/srep25279>.
- Abd El-Wakeil KF. Effects of terrestrial isopods (Crustacea: Oniscidea) on leaf litter decomposition processes. *J Basic Appl Zool*. 2015;69:10–6.
- Jambu P, Juchault P, Mocquard JP. Étude expérimentale de la contribution du crustacé isopode *Oniscus asellus* à la transformation des litières forestières sous chêne sessile. *Pedobiologia*. 1988;32:147–56.
- Mocquard JP, Juchault P, Jambu P, Fustec E, Lebourg B. Essai d'évaluation du rôle des crustacés oniscoides dans la transformation des litières végétales dans une forêt feuillue de l'ouest de la France. *Rev Écol Biol Sol*. 1987;3:311–27.
- Zimmer M, Kautz G, Topp W. Do woodlice and earthworms interact synergistically in leaf litter decomposition? *Funct Ecol*. 2005;19:7–16.
- Hättenschwiler S, Tiunov AV, Scheu S. Biodiversity and litter decomposition in terrestrial ecosystems. *Annu Rev Ecol Evol Syst*. 2005;36:191–218.
- Hassall M, Turner JG, Rands MRW. Effects of terrestrial isopods on the decomposition of woodland leaf litter. *Oecologia*. 1987;72:597–604.
- Špaldoňová A, Frouz J. The role of *Armadillidium vulgare* (Isopoda: Oniscidea) in litter decomposition and soil organic matter stabilization. *Appl Soil Ecol*. 2014;83:186–92.
- Hassall M. The functional morphology of the mouthparts and foregut in the terrestrial isopod *Philoscia muscorum*. *Crustaceana*. 1977;33:225–36.
- Jia Y, Lv Y, Kong X, Jia X, Tian K, Du J, et al. Insight into the indirect function of isopods in litter decomposition in mixed subtropical forests in China. *Appl Soil Ecol*. 2015;86:174–81.
- Broly P, Deville P, Mailet S. The origin of terrestrial isopods (Crustacea: Isopoda: Oniscidea). *Evol Ecol*. 2013;27:461–76.
- Edney EB. Transition from water to land in isopod crustaceans. *Am Zool*. 1968;8:309–26.
- Warburg MR. Isopods and their terrestrial environment. In: *Advances in ecological research*; Elsevier; 1987. p. 187–242. [https://doi.org/10.1016/S0065-2504\(08\)60246-9](https://doi.org/10.1016/S0065-2504(08)60246-9).
- Zimmer M. Nutrition in terrestrial isopods (Isopoda: Oniscidea): an evolutionary-ecological approach. *Biol Rev*. 2002;77:455–93.
- Zimmer M, Danko JP, Pennings SC, Danford AR, Carefoot TH, Ziegler A, et al. Cellulose digestion and phenol oxidation in coastal isopods (Crustacea: Isopoda). *Mar Biol*. 2002;140:1207–13.
- Hames CAC, Hopkin SP. The structure and function of the digestive system of terrestrial isopods. *J Zool*. 1989;217:226–36.
- Hassall M, Jennings JB. Adaptive features of gut structure and digestive physiology in the terrestrial isopod *Philoscia muscorum*. *Biol Bull*. 1975;149:348–64.
- Zimmer M, Bartholmé S. Bacterial endosymbionts in *Asellus aquaticus* (Isopoda) and *Gammarus pulex* (Amphipoda) and their contribution to digestion. *Limnol Oceanogr*. 2003;48:2208–13.
- Graça MAS, Maltby L, Calow P. Importance of fungi in the diet of *Gammarus pulex* and *Asellus aquaticus* I: feeding strategies. *Oecologia*. 1993;93:139–44.
- Kern M, McGeehan JE, Streeter SD, Martin RNA, Besser K, Elias L, et al. Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proc Natl Acad Sci*. 2013;110:10189–94.
- Kostanjšek R, Milatović M, Štrus J. Endogenous origin of endo-β-1,4-glucanase in common woodlouse *Porcellio scaber* (Crustacea, Isopoda). *J Comp Physiol B*. 2010;180:1143–53.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42:D490–5.
- Bredon M, Dittmer J, Noël C, Moumen B, Bouchon D. Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate. *Microbiome*. 2018;6. <https://doi.org/10.1186/s40168-018-0536-y>.
- King AJ, Cragg SM, Li Y, Dymond J, Guille MJ, Bowles DJ, et al. Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci*. 2010;107:5345–50.
- Besser K, Malyon GP, Eborall WS, Paro da Cunha G, Filgueiras JG, Dowle A, et al. Hemocyanin facilitates lignocellulose digestion by wood-boring marine crustaceans. *Nat Commun*. 2018;9. <https://doi.org/10.1038/s41467-018-07575-2>.
- Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, et al. Lignocellulose degradation mechanisms across the tree of life. *Curr Opin Chem Biol*. 2015;29:108–19.
- Brune A. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*. 2014;12:168–80.
- Ni J, Tokuda G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv*. 2013;31:838–50.
- Rizzi A, Crotti E, Borruso L, Jucker C, Lupi D, Colombo M, et al. Characterization of the bacterial community associated with larvae and adults of *Anoplophora chinensis* collected in Italy by culture and culture-independent methods. *Biomed Res Int*. 2013;2013:1–12.
- Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, et al. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. *PLoS One*. 2013;8:e73827.
- Bouchon D, Zimmer M, Dittmer J. The terrestrial isopod microbiome: an all-in-one toolbox for animal–microbe interactions of ecological relevance. *Front Microbiol*. 2016;7. <https://doi.org/10.3389/fmicb.2016.01472>.
- Kostanjšek R, Lapanje A, Rupnik M, Štrus J, Drobne D, Avguštin G. Anaerobic bacteria in the gut of terrestrial isopod crustacean *Porcellio scaber*. *Folia Microbiol (Praha)*. 2004;49:179–82.
- Kostanjšek R, Štrus J, Lapanje A, Avguštin G, Rupnik M, Drobne D. Intestinal microbiota of terrestrial isopods. In: *Intest Microorg Termit Invertebr.*: Springer-Verlag Berlin; 2006. p. 115–31.
- Wang Y, Brune A, Zimmer M. Bacterial symbionts in the hepatopancreas of isopods: diversity and environmental transmission: bacterial symbionts in isopods. *FEMS Microbiol Ecol*. 2007;61:141–52.
- Zimmer M, Topp W. Microorganisms and cellulose digestion in the gut of the woodlouse *Porcellio scaber*. *J Chem Ecol*. 1998;24:1397–408.
- Han C, Li Q, Li X, Zhang Z, Huang J. *De novo* assembly, characterization and annotation for the transcriptome of *Sphaeroma terebrans* and microsatellite marker discovery. *Genes Genomics*. 2018;40:167–76.
- Zimmer M. The role of animal–microbe interactions in isopod ecology and evolution. *Acta Biol Benrodis*. 2006;13:127–68.
- Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, et al. An ancient family of lytic polysaccharide monoxygenases with roles in arthropod development and biomass digestion. *Nat Commun*. 2018;9. <https://doi.org/10.1038/s41467-018-03142-x>.
- Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2013;14:274.
- Warnecke F, Luginbühl P, Ivanova N, Ghasseman M, Richardson TH, Stege JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*. 2007;450:560–5.

43. Zhang W, Zhang X, Li K, Wang C, Cai L, Zhuang W, et al. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. *Mycology*. 2018;9:176–88.
44. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science*. 2012;336:1715–9.
45. Davison A, Blaxter M. Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol*. 2005;22:1273–84.
46. Lo N, Watanabe H, Sugimura M. Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals. *Proc R Soc B Biol Sci*. 2003;270(Suppl_1):S69–72.
47. Lins LSF, Ho SYW, Lo N. An evolutionary timescale for terrestrial isopods and a lack of molecular support for the monophyly of Oniscidea (Crustacea: Isopoda). *Org Divers Evol*. 2017;17:813–20.
48. Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*. 2007;315:804–7.
49. Zimmer M, Danko JP, Pennings SC, Danford AR, Ziegler A, Uglow RF, et al. Hepatopancreatic endosymbionts in coastal isopods (Crustacea: Isopoda), and their contribution to digestion. *Mar Biol*. 2001;138:955–63.
50. Tan MH, Gan HM, Gan HY, Lee YP, Croft LJ, Schultz MB, et al. First comprehensive multi-tissue transcriptome of *Cherax quadricarinatus* (Decapoda: Parastacidae) reveals unexpected diversity of endogenous cellulase. *Org Divers Evol*. 2016;16:185–200.
51. Gray M, Linton SM, Allardyce BJ. cDNA sequences of GHF9 endo- β -1,4-glucanases in terrestrial Crustacea. *Gene*. 2018;642:408–22.
52. Rytioja J, Hildén K, Yuzon J, Hatakka A, de Vries RP, Mäkelä MR. Plant-polysaccharide-degrading enzymes from Basidiomycetes. *Microbiol Mol Biol Rev*. 2014;78:614–49.
53. Kao D, Lai AG, Stamatakis E, Rosic S, Konstantinides N, Jarvis E, et al. The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife*. 2016;5:e20062.
54. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, et al. The Ecoresponsive genome of *Daphnia pulex*. *Science*. 2011;331:555–61.
55. Mayer WE, Schuster LN, Bartelmes G, Dieterich C, Sommer RJ. Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evol Biol*. 2011;11. <https://doi.org/10.1186/1471-2148-11-13>.
56. Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*. *FEBS Lett*. 2004;572:201–5.
57. Danchin EJ, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci*. 2010;107:17651–6.
58. Eyun S, Wang H, Pauchet Y, French-Constant RH, Benson AK, Valencia-Jiménez A, et al. Molecular evolution of glycoside hydrolase genes in the Western corn rootworm (*Diabrotica virgifera virgifera*). *PLoS One*. 2014;9:e94052.
59. Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci*. 2012;109:4197–202.
60. Ioannidis P, Lu Y, Kumar N, Creasy T, Daugherty S, Chibucos MC, et al. Rapid transcriptome sequencing of an invasive pest, the brown marmorated stink bug *Halyomorpha halys*. *BMC Genomics*. 2014;15:738.
61. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from diverse Bacteria to an insect genome enables a tripartite nested mealybug Symbiosis. *Cell*. 2013;153:1567–78.
62. Janusz G, Pawlik A, Sulej J, Świdarska-Burek U, Jarosz-Wilkolazka A, Paszczyński A. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol Rev*. 2017;41:941–62.
63. Zimmer M. The fate and effects of ingested hydrolyzable tannins in *Porcellio scaber*. *J Chem Ecol*. 1999;25:611–28.
64. Zimmer M, Topp W. Nutritional biology of terrestrial isopods (Isopoda: Oniscidea): copper revisited. *Isr J Zool*. 1998;44:453–62.
65. Sützl L, Laurent CVFP, Abrera AT, Schütz G, Ludwig R, Haltrich D. Multiplicity of enzymatic functions in the CAZy AAG family. *Appl Microbiol Biotechnol*. 2018;102:2477–92.
66. Sun W, Shen Y-H, Yang W-J, Cao Y-F, Xiang Z-H, Zhang Z. Expansion of the silkworm GMC oxidoreductase genes is associated with immunity. *Insect Biochem Mol Biol*. 2012;42:935–45.
67. Iida K, Cox-Foster DL, Yang X, Ko W-Y, Cavener DR. Expansion and evolution of insect GMC oxidoreductases. *BMC Evol Biol*. 2007;7:75.
68. Yanagisawa M, Kawai S, Murata K. Strategies for the production of high concentrations of bioethanol from seaweeds: production of high concentrations of bioethanol from seaweeds. *Bioengineered*. 2013;4:224–35.
69. Si A, Bellwood O, Alexander CG. Evidence for filter-feeding by the wood-boring isopod, *Sphaeroma terebrans* (Crustacea: Peracarida). *J Zool*. 2002;256:463–71.
70. Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. *Microbiome*. 2018;6. <https://doi.org/10.1186/s40168-018-0457-9>.
71. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
72. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA – a practical iterative de Bruijn graph De novo assembler. In: Berger B, editor. Research in computational molecular biology. Berlin: Springer Berlin Heidelberg; 2010. p. 426–40. https://doi.org/10.1007/978-3-642-12683-3_28.
73. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
74. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2.
75. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40:W445–51.
76. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013;41:e121.
77. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
78. Huson DH, Beier S, Flade I, Górski A, El-Hadidi M, Mitra S, et al. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol*. 2016;12:e1004957.
79. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
80. Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. *BMC Bioinformatics*. 2017;18. <https://doi.org/10.1186/s12859-017-1625-9>.
81. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res*. 2015;43:W566–70.
82. Emms DM, Kelly S. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences; 2018. <https://doi.org/10.1101/466201>.
83. Emms D, Kelly S. STAG: species tree inference from all genes; 2018. <https://doi.org/10.1101/267914>.
84. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006;22:1269–71.
85. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279–85.
86. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
87. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007;56:564–77.
88. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
89. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
90. Hoang DT, Chernomor O. UFBoot2: improving the ultrafast bootstrap approximation; 2017. p. 5.
91. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44:W242–5.
92. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol*. 2018;35:773–7.
93. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. *Mol Biol Evol*. 2015;32:1365–71.

94. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005; 22:1208–22.
95. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. *Mol Biol Evol.* 2013;30:1196–205.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

