

RESEARCH

Open Access



# Isopod holobionts as promising models for lignocellulose degradation

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

## Abstract

**Background:** Isopods have colonized all environments, partly thanks to their ability to decompose the organic matter. Their enzymatic repertoire, as well as the one of their associated microbiota, has contributed to their colonization success. Together, these holobionts have evolved several interesting life history traits to degrade the plant cell walls, mainly composed of lignocellulose. It has been shown that terrestrial isopods achieve lignocellulose degradation thanks to numerous and diverse CAZymes provided by both the host and its microbiota. Nevertheless, the strategies for lignocellulose degradation seem more diversified in isopods, in particular in aquatic species which are the least studied. Isopods could be an interesting source of valuable enzymes for biotechnological industries of biomass conversion.

**Results:** To provide new features on the lignocellulose degradation in isopod holobionts, shotgun sequencing of 36 metagenomes of digestive and non-digestive tissues was performed from several populations of four aquatic and terrestrial isopod species. Combined to the 15 metagenomes of an additional species from our previous study, as well as the host transcriptomes, this large dataset allowed us to identify the CAZymes in both the host and the associated microbial communities. Analyses revealed the dominance of *Bacteroidetes* and *Proteobacteria* in the five species, covering 36% and 56% of the total bacterial community, respectively. The identification of CAZymes and new enzymatic systems for lignocellulose degradation, such as PULs, cellulosomes and LPMOs, highlights the richness of the strategies used by the isopods and their associated microbiota.

**Conclusions:** Altogether, our results show that the isopod holobionts are promising models to study lignocellulose degradation. These models can provide new enzymes and relevant lignocellulose-degrading bacteria strains for the biotechnological industries of biomass conversion.

**Keywords:** Lignocellulose, CAZymes, Shotgun metagenomics, Microbiota, Holobiont, Transcriptomics, Isopods

## Background

Microbiota shapes living organisms through complex interactions. Under changing environmental conditions, it may rapidly evolve and influence on the host adaptation and evolution. The microbiota is known to act on animal development, as well as animal health and evolution [1]. Since the “microbiome revolution” of the last

10 years [2], many studies have highlighted the impact of microbiota on the fitness of the host. This change in our vision of organisms led to the recent introduction of the holobiont concept. This concept considers a holobiont as a combination of a host and its associated microbial community, including bacteria, viruses and cellular organisms [3–5]. As a result, a holobiont is an assemblage of species that are metabolically interdependent. Interaction patterns in these systems shape the holobiont’s composition [6], and conversely, bionts (i.e. members of the holobiont) can be agents of developmental plasticity that facilitate the evolution of new phenotypes in animals.

\*Correspondence: didier.bouchon@univ-poitiers.fr  
Laboratoire Ecologie et Biologie des Interactions-UMR CNRS 7267,  
Ecologie et Biologie des Interactions-Bâtiment B8-B35, Université de  
Poitiers, 5 rue Albert Turpin, TSA 51 106, 86073 Poitiers Cedex 9, France



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Herbivory, and more specifically lignocellulose degradation, is one of these processes that evolved through symbiont acquisition in many animals [7]. In the context of global climate change, lignocellulose is also an important renewable and sustainable source to produce biofuels and other bioproducts [8]. It seems to be the best alternative to fossil fuels, thus attracting attention from researchers and industrials worldwide. Bacteria and fungi have traditionally been used for research of lignocellulose-degrading enzymes due to their important role in the decomposition of organic matter in ecosystems [9–11]. For most animals, the degradation of lignocellulose involves the cooperation of many bionts to fully achieve its deconstruction [12]. It requires a large number of enzymes that are classified in Carbohydrate Active EnZymes (also called CAZymes) families [13]. CAZymes act on lignocellulose like an enzymatic cocktail; they complement each other and work in synergy to degrade each component of the lignocellulose. Lignocellulose is mainly composed of cellulose, hemicellulose and lignin [14]. CAZymes are thus classified in three classes, depending on the targeted substrate: cellulases, hemicellulases and lignin-modifying enzymes (abbreviated hereafter LME). These three types of enzymes are classified in several CAZy families: the majority of cellulases belong to glycoside hydrolases (GH), the hemicellulases to carbohydrate esterases (CE) and to GHs, and the LME are all classified in auxiliary activities (AA). In addition, recent studies have shown the existence of oxidative cellulases classified in AA families [15]. They break down cellulose with oxidative processes, contrary to the cellulases classified as GHs that hydrolyze cellulose [16, 17]. However, the recalcitrance of lignocellulose, as well as the strong demand of novel enzymes by the industry, imply to explore new models for lignocellulose degradation. Because of the insufficient quantities of enzymes produced by fungi, new research now focuses on bacterial lignocellulose-degrading CAZymes [11]. Expanding bacteria models to the holobiont might enable us to find new strategies for lignocellulose degradation, and thus to respond to the lack of resources for the energetic and chemical industries.

In this context, isopods could be very interesting models to study the lignocellulose degradation in the light of the holobiont concept. Strategies for lignocellulose degradation are different between terrestrial and aquatic isopods [18]. Aquatic isopods have developed specific strategies to degrade the lignocellulose, and for some of them without the help of microbiota [19–21]. The best example is *Limnoria quadripunctata*, a marine wood-boring isopod which feeds on the cellulose without any help from the microbes [21]. Remarkably, whereas hemocyanins conventionally stand as respiratory proteins,

those secreted in the hindgut of *L. quadripunctata* modify the lignin and thus enhance the digestibility of cellulose [19]. On the other hand, terrestrial isopods shelter rich and diverse microbial communities in all their tissues [22–24], implying multiple interactions in the holobiont. These communities enable an efficient digestion of the lignocellulose thanks to a complementarity between their CAZome (i.e. CAZyme repertoire) and that of their host [18, 25, 26]. Furthermore, since the land conquest, the CAZomes of terrestrial isopods have been enriched through several gene duplications and horizontal transfers [18]. Strategies for lignocellulose degradation in isopod therefore depend in part on the interactions within the holobiont. Moreover, it has been estimated that herbivory arose independently three times in isopods [27] and promoted Crustacea diversification [28]. As a result, we might suppose that much remains unknown on lignocellulose degradation strategies in isopods.

The main purpose of this study is to provide new features on lignocellulose degradation in the isopod holobionts. To this end, shotgun sequencing of 36 metagenomes was performed from digestive and non-digestive tissues of one freshwater and three terrestrial isopod species from several populations. In addition, we used 15 shotgun metagenomes of our previous study on the pill bug *Armadillidium vulgare* [25]. Combined with the host transcriptomic data, this large dataset enabled us to identify the CAZome of both host and microbiota, and microbial taxa associated with lignocellulose degradation. The comparison of different CAZomes, as well as the identification of CAZymes, PULs (“Polysaccharide Utilization Loci”) and cellulosomes in isopods, highlighted the diversity of strategies for this process in isopods. Altogether, these results show that isopod holobionts are promising models to study lignocellulose degradation and to potentially discover new lignocellulose-degrading CAZymes.

## Results

### Quality of metagenome and transcriptome assemblies

To build the CAZomes of the isopod holobionts, 51 metagenomic samples from digestive and non-digestive tissues of five isopod species were processed from 36 new datasets, along with 15 datasets from our previous study [25]. Samples from different origins enabled us to compare CAZomes of different populations for a single species. These samples represented a total of 5.7 billion reads, assembled into 25 million contigs including 4.7 million contigs > 1 kb length (Additional file 1).

Meanwhile, host transcriptomes were obtained from whole individuals from several populations (for more details, see [18]). Transcriptome assemblies resulted in 40,916 from 143,383 transcripts depending on the

species. Assemblies showed an N50 from 1096 bp to 1523 bp depending on the transcriptome (Additional file 2). They displayed a good completeness since more than 95% of the complete genes from the arthropod core genome were present in the *A. vulgare*, *P. dilatatus dilatatus*, *P. dilatatus petiti* and *P. pruinus* assemblies, and 83% of these genes were present in the *A. aquaticus* assembly (Additional file 2).

### CAZyme identification in isopod holobionts

Both contigs and transcripts from metagenomes and host transcriptomes were subjected to the CAZy database (<http://www.cazy.org>) to identify CAZomes in isopod holobionts. The use of dbCAN2 for the identification of CAZymes resulted in more stringent criteria than in our previous studies [18, 25] for the identification of CAZymes. In total, 15,834 CAZymes were identified distributed among 201 CAZy families of which 27 were specific to the hosts, 136 were found only in microbiota, and 36 were present in both (Additional files 3, 4, 5). Of these CAZymes, 12,916 belonged to the metagenomes (distributed among 174 families) and 2918 belonged to the host transcriptomes (distributed among 63 families). Eighty-four GH families were identified in the CAZomes of isopod holobionts; they represented the most abundant and diversified families of those CAZomes. GHs are a prominent group of enzymes that hydrolyze the glycosidic bonds, most of cellulases and hemicellulases belong to these families. In the metagenomes, GH13 was the most frequently occurring GH family (799 modules identified in 82.4% of the samples), while GH18 represented the most abundant GH family in transcriptomes (315 modules identified in all samples) (Additional file 3). Then, GTs families were the second most abundant ones with 55 members, of which GT2 was the most abundant GT family in metagenomes (1493 modules identified in 75% of the samples) and GT1 the most abundant in transcriptomes (333 modules identified in all samples). As regards other families, 27 CBMs, 13 CEs, 13 PLs and 9 AAs were identified.

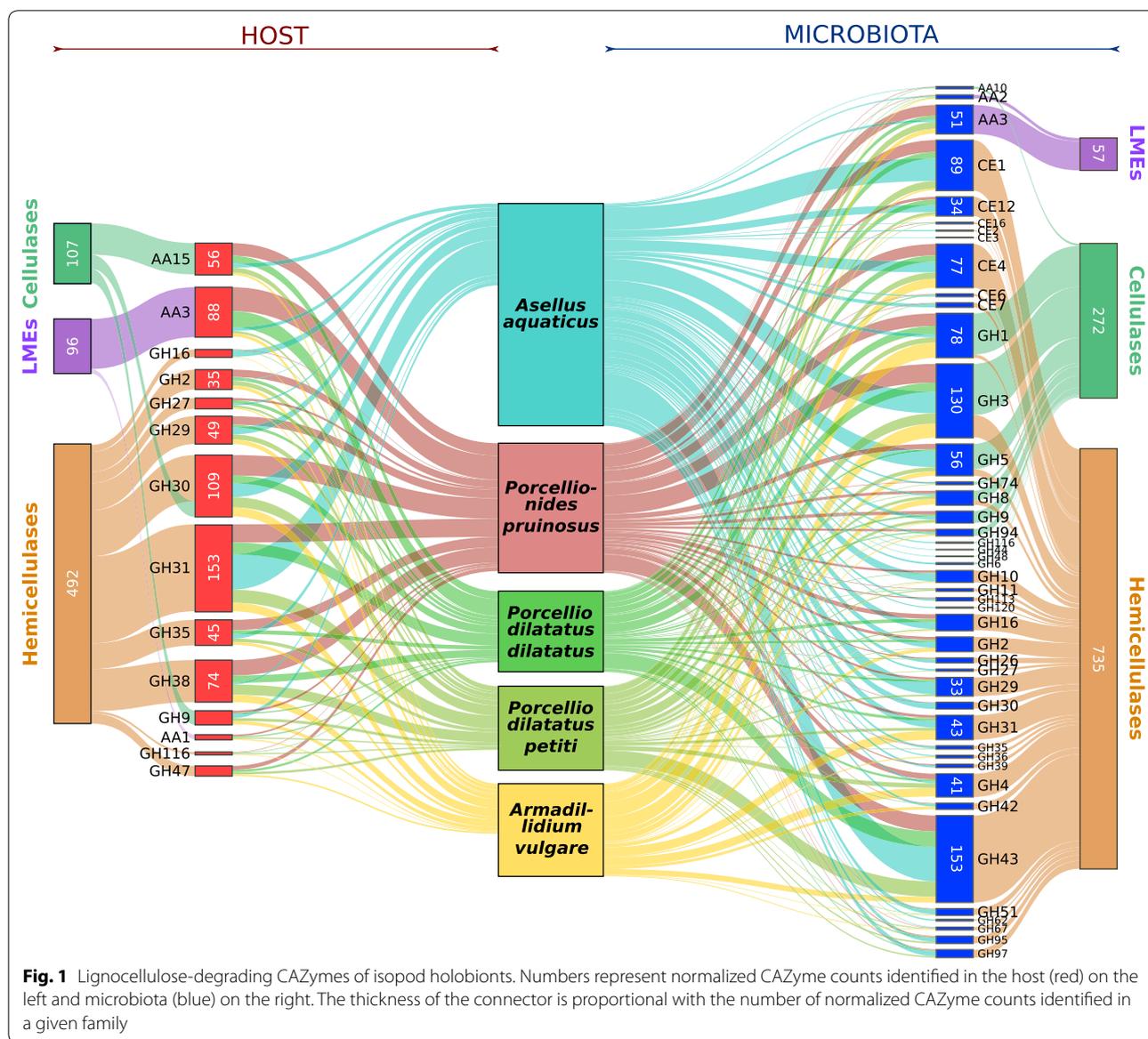
### Lignocellulose-degrading CAZymes

Selected CAZymes likely to contribute to the lignocellulose degradation were then examined in depth. Focusing on digestive enzymes in isopod holobionts, only metagenomic samples of caeca and hindgut were considered for the subsequent analyses. In total, 44 lignocellulose-degrading CAZyme families representing 3140 modules were predicted in the metagenomes (Additional file 3), distributed among 33 GH families, eight CE families and three AA families (Fig. 1). Enzymatic activities of 1313 (41.8%) of these modules were predicted by Hotpep (Additional file 6). Among them,

1074 (81.8%) could have a lignocellulosic enzyme activity and 239 (18.2%) potentially act on other substrates (e.g. pectin, chitin).

Most of lignocellulose-degrading CAZymes were identified in the hindgut, especially the LMEs that were absent from the caeca in all species, except *P. pruinus* (Fig. 2). In all the species, hemicellulases known to degrade a broad range of substrates like xylan, mannan and xyloglucan were the most abundant lignocellulose-degrading CAZyme in the metagenomes with 36 families in total (Fig. 1). This is especially the case for *A. aquaticus* metagenomes in which hemicellulases represent 76% of the identified lignocellulose-degrading CAZyme modules. Concerning cellulases, 12 families were identified in the metagenomes, including seven endocellulases (GH5, GH6, GH8, GH9, GH44, GH48, GH74), three beta-glucosidases (GH1, GH3, GH116), one cellobiose phosphorylase (GH94) and one LPMO (AA10) (Fig. 1). Note that four CAZyme families (GH5, GH8, GH1 and GH3) contain both cellulases and hemicellulases (Fig. 1 and Additional file 6). Finally, two families of LMEs were identified in the metagenomes: one family of laccases (AA1) and one family of cellobiose dehydrogenases (AA3).

Among the 2918 CAZymes belonging to the host transcriptomes, 987 could act as lignocellulose-degrading CAZymes (Additional file 3). The enzyme activity of 567 (57.45%) of those CAZymes was predicted using Hotpep, 555 (97.9%) of which could degrade the lignocellulose (Additional file 6). They were distributed among 14 CAZy families including 11 GHs and three AAs (Fig. 1). Host lignocellulose-degrading CAZymes were thus three times less diversified and abundant than in the microbiota, and only four were specific: GH47, GH38, AA15, and AA1. Three cellulases were identified in isopod hosts: two cellulases widespread in crustaceans, belonging to the GH9 and GH30 families [20, 29, 30], and an LPMO, belonging to the AA15 family, that was recently identified in arthropods [17, 18]. Concerning LMEs, a few laccases belonging to the AA1 family were identified in terrestrial isopods but not in the aquatic species, and some AA3s were found in the five species (Fig. 1). Finally, just like for the metagenomes, hemicellulases were the most diversified CAZymes in the host transcriptomes with 11 families identified. Compared to the terrestrial species, the microbiota of the freshwater isopod *A. aquaticus* seems to contribute more than the host to lignocellulose degradation (Fig. 1). Indeed, only 11 lignocellulose-degrading CAZymes were identified from the transcriptome, whereas 43 lignocellulose-degrading CAZymes were identified from the metagenome, which is 1.5 times more than in the metagenomes of terrestrial species (containing 28 lignocellulose-degrading CAZymes on average).

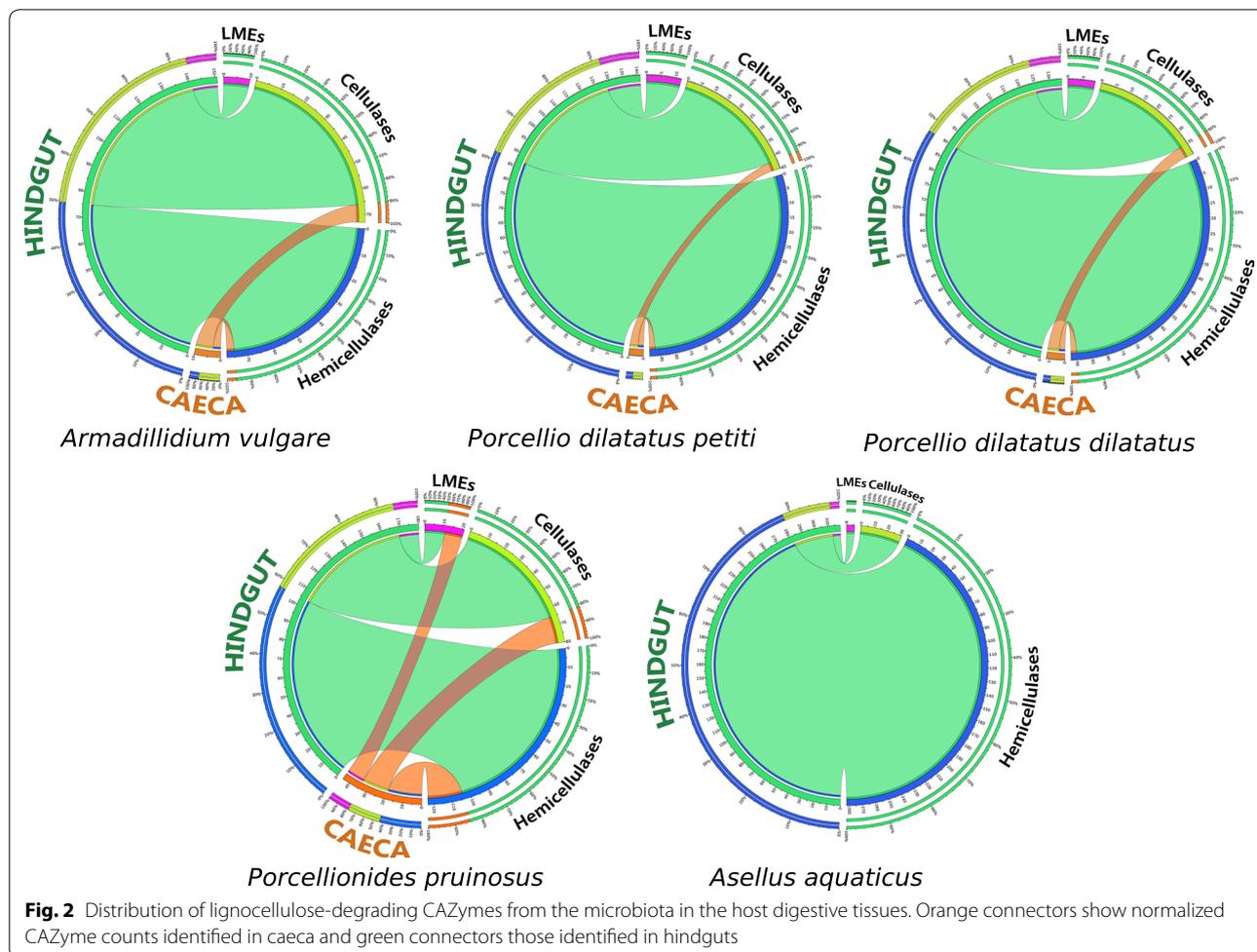


### Taxonomic origin of the CAZymes from microbiota

To identify the microbial communities associated to lignocellulose degradation, similarity searches of the predicted CAZymes from the metagenomes were performed against the NCBI Non-Redundant Protein database. Proteobacteria and Bacteroidetes were the most represented bacterial phyla, accounting for 56% and 36% of the identified CAZymes, respectively (Fig. 3a). The microbial communities were rich and highly diversified in all host species. Indeed, 41 bacterial orders were found in all microbiota (Fig. 3b). The taxonomic origin of the microbial CAZymes was highly different from one host to another (Figs. 3, 4), without apparent sex effect except for *P. pruinosus* (Fig. 4a).

Moreover, the microbial CAZymes were shaped by the environment, as for a given species these communities differed according to the host origin (field or laboratory) (Fig. 4). Compared to terrestrial isopods, the microbiota of *A. aquaticus* included more Bacteroidetes and less Proteobacteria (Fig. 3a). Furthermore, there was a high percentage of non-identified bacteria, notably for the communities that encode GH families in the *A. aquaticus* laboratory lineage, where more than half of bacterial families were unknown (Fig. 3b).

Flavobacteriales (Bacteroidetes), Micrococcales (Actinobacteria) and Burkholderiales (Proteobacteria) were the largest contributors to lignocellulose-degrading CAZymes, especially GH, CE and AA families, in both



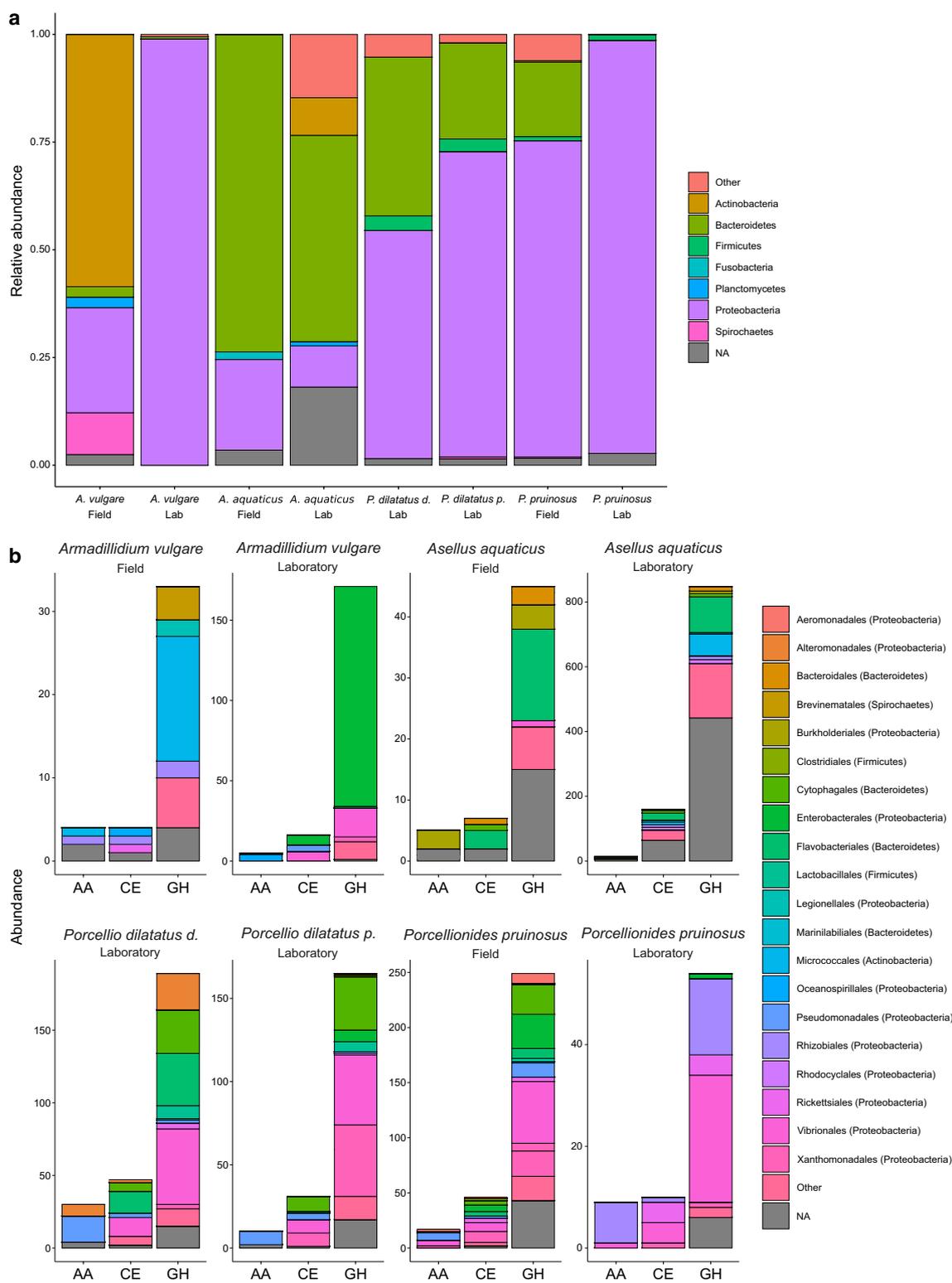
populations of *A. aquaticus* (Fig. 3). In the terrestrial host species, Vibrionales (Proteobacteria) encoded many lignocellulose-degrading CAZymes belonging to GH families in all hosts, as well as AA and CE families in some of them. Cytophagales (Bacteroidetes) were abundant in *P. pruinosus* (field population), *P. dilatatus dilatatus* and *P. dilatatus petiti*. Likewise, Rhizobiales (Proteobacteria) encoded the most part of lignocellulose-degrading CAZymes in *P. pruinosus* originating from laboratory. Finally, among other bacterial families, Xanthomonadales (Proteobacteria), Enterobacteriales (Proteobacteria) and Flavobacteriales (Bacteroidetes) were also important contributors to lignocellulose degradation in some terrestrial isopod species.

#### Identification of PULs and cellulosomes in the microbiota

Potential PULs were screened by searching within contigs for both gene markers of PULs: *SusC* and *SusD*. The *SusC* or/and the *SusD* conserved domains were identified in 5089 contigs. Among them, only 37 contigs encoded one sequential pair of *susC* and *susD* (Additional file 7).

They were found in all species except *A. vulgare* and the laboratory lineage of *P. pruinosus*. Unsurprisingly, all these contigs were assigned to Bacteroidetes species. In the metagenomes of *A. aquaticus*, three contigs harboring *Sus* genes were of interest as they also encoded CAZymes which might be involved in lignocellulose degradation (Fig. 5). Indeed, they could act in the breakdown of hemicellulosic substrates thanks to their GH3, GH16 and GH43 enzymes. These PULs exhibited a gene organization different from those referenced in the PUL DataBase [31].

Potential cellulosomes were predicted by searching cohesin or dockerin modules within contigs. In total, 835 dockerin modules and 65 cohesin modules were identified in 874 contigs of the metagenomes (Additional files 8, 9). In addition, 163 contigs harbored an SLH domain (Additional files 8, 9), an anchoring module that helps to bind the cellulosome to the cell surface. The presence of a dockerin domain in 30 CAZyme genes provided evidence for active cellulosomes in microbiota (Table 1). Several of those CAZymes are known to deconstruct



**Fig. 3** Compositions of the bacterial community associated with the lignocellulose degradation at the level of the phylum (a) and of the order (b) according to CAZy classes

(See figure on next page.)

**Fig. 4** Comparative analysis of bacterial communities associated with the lignocellulose degradation in isopods, considering (a) or not (b) the effect of sex. **a** A principal component analysis (PCA) plot of the bacterial counts, at the family level, that characterize the bacterial community from each studied species depending of its origin and sex (51.2% of the information was extracted from the two principal components PC1 and PC2). Each dot represents the taxonomic composition of a metagenome and each color represents the host species and its origin (laboratory or field). **b** Phylogenetic tree of the bacterial communities of the studied isopods. All branches are drawn to scale as indicated by the scale bar

lignocellulose: CE1, CE3, GH2, GH3, GH5, GH26, GH35, GH43 and GH44. In addition, one GT (GT102), three CBMs (CBM32, CBM35, CBM38), one PL (PL9), nine other GHs (GH18, GH33, GH50, GH93, GH99, GH108, GH130, GH135, GH136) and one other CE (CE10) were also associated with dockerin domains. The taxonomic assignment of these contigs showed that potential cellulosomes were carried out by bacteria belonging to Bacteroidetes, Planctomycetes, Armatimonadetes, Rhizobiales and several unknown bacteria (Table 1).

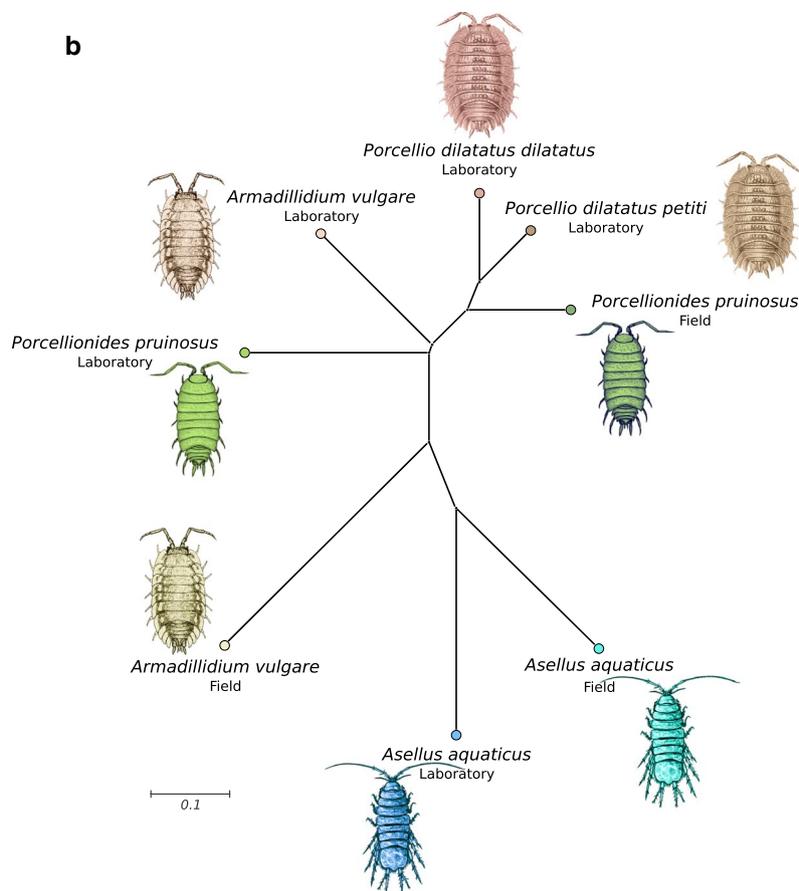
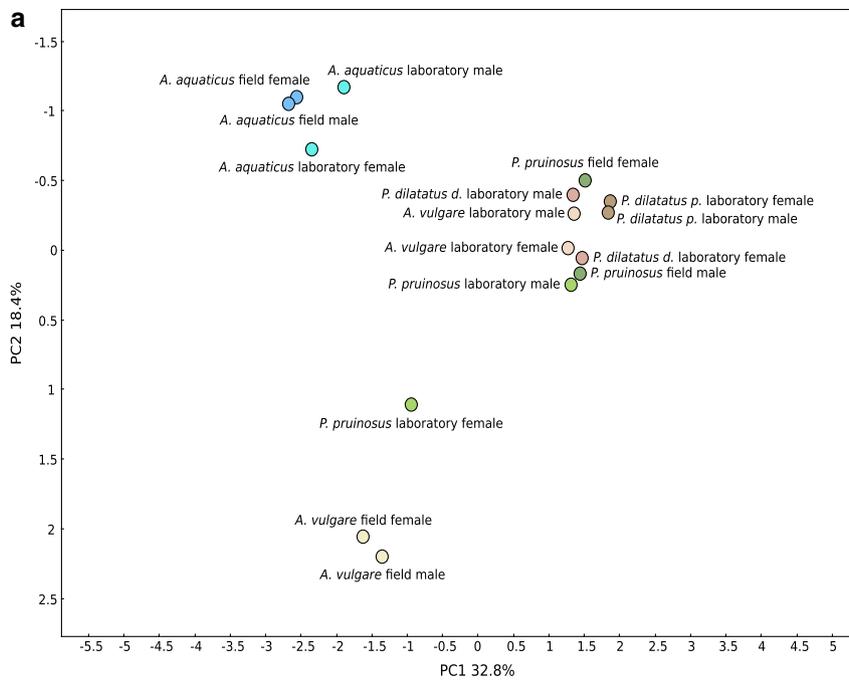
## Discussion

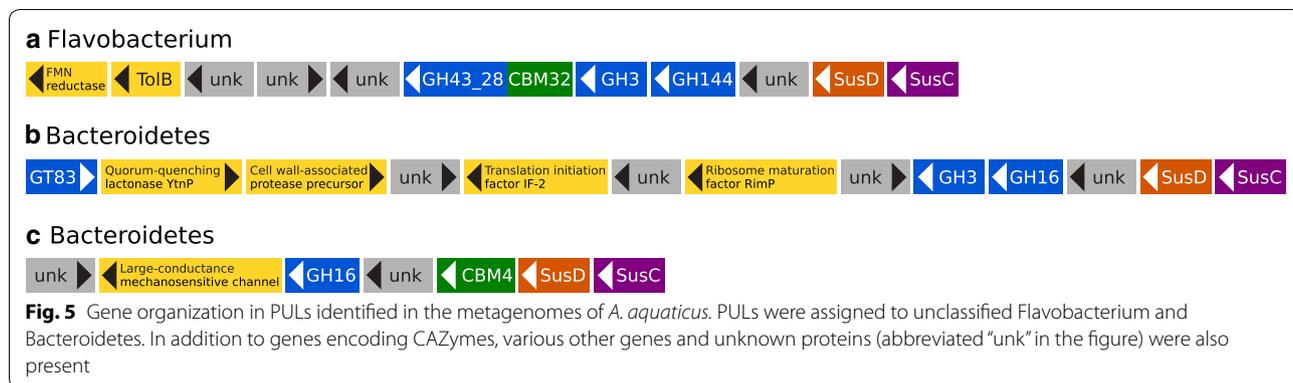
We investigated the repertoire for lignocellulose degradation on the scale of the holobiont in five isopod species, aquatic and terrestrial, highlighting that host and microbiota complement their CAZomes to achieve effective lignocellulose deconstruction. Despite using more stringent criteria than in our previous studies [18, 25], we found a great diversity of CAZymes in all isopod holobionts. We also showed that highly different host-associated bacterial communities occurred in the terrestrial species and even within species, revealing a functional redundancy in lignocellulose degradation and as regards complementarity with the host's repertoire. In addition, we highlighted the potential involvement of microbial cellulosomes, as well as the presence of PUL systems in isopods. Isopods therefore represent an unexploited wealth of lignocellulose-degrading enzymes and natural nanomachines.

The microbiota had a prominent weight in the lignocellulose-degradation repertoire of the five isopod holobionts: the microbial lignocellulose-degrading CAZymes exceeded by three times those of the hosts. A large part was not represented in the hosts: accordingly, the diverse microbial CAZymes might complete the CAZome of the host for efficient lignocellulose degradation. Considering that host CAZymes are primarily produced in the caeca [25], the production of CAZymes in the hindgut by microbiota contributes to sequential biomass degradation in the digestive organs. Spatially, the digestion of lignocellulose is therefore performed along the different parts of the digestive tract thanks to the successive intervention of host and microbial enzymes from the foregut where the caeca (or digestive glands) open to the posterior part of the hindgut [26, 32, 33].

Hemicellulases are the most numerous lignocellulose-degrading CAZymes in isopod holobionts. Since hemicellulose composition varies from one plant to another, and even from one plant tissue to another [16], organisms require large repertoires of hemicellulases to be able to degrade these complex structures. As decomposers, isopods have to deal with a wide variety of foods. The diversity of hemicellulases we recorded in both host and microbiota might be an adaptive response of isopod holobionts to the complex composition of hemicellulose. Several types of cellulases belonging to endocellulases and beta-glucosidases were also identified in both host and microbiota. In contrast, no exocellulases were found in the five studied isopods. Whereas they are required by fungi to achieve the cellulose degradation [34], they are rarely found in animals, among which a few crustaceans, including five isopod species not corresponding to those presently studied [13, 18, 30, 35, 36]. Contrary to termites where exocellulases are provided by their symbionts [37–39], isopods may degrade the cellulose thanks to the abundance and diversity of other kinds of cellulases provided by their microbiota. Furthermore, we have identified oxidative cellulases (LPMO) belonging to AA15 family in host transcriptomes and LPMO belonging to AA10 in the metagenomes of *A. aquaticus*, *P. pruinosus* and *P. dilatatus p.*, suggesting the existence of alternative strategies for cellulose degradation in isopod holobionts. Finally, isopod hosts might be key players of lignin modification for a better exploitation of hemicellulose and cellulose. Numerous cellobiose dehydrogenases (CDH) belonging to the AA3 family and laccases belonging to AA1 family were found in the four terrestrial host species. Laccases are among the most important LME in wood-destroying microorganisms [40, 41] but CDHs are not widespread in arthropods and are even absent in insects [18]. In addition, in the five isopod species, the microbiota provides manganese peroxidases belonging to the AA2 family, as well as some AA3 enzymes. Once again, the host–microbiota cooperation could permit isopod to efficiently breakdown the lignin allowing access to other lignocellulose components.

It is interesting to note that host–microbiota contribution to lignocellulose degradation appeared different in the freshwater isopod *A. aquaticus* compared to the terrestrial ones. The host CAZome of *A. aquaticus* is less expanded than those of terrestrial species [18].





**Table 1** CAZymes containing dockerin domains

Host	Origin	Gender	Tissue	Assignment	CAZymes
<i>A. aquaticus</i>	Field	Female	Nd. <sup>a</sup> tissues	Bacteria	GH35
<i>A. aquaticus</i>	Field	Female	Nd. <sup>a</sup> tissues	Bacteroidetes	GH130
<i>A. aquaticus</i>	Field	Female	Nd. <sup>a</sup> tissues	Cytophagales	GH35
<i>A. aquaticus</i>	Field	Male	Nd. <sup>a</sup> tissues	Bacteria	GH130
<i>A. aquaticus</i>	Field	Male	Nd. <sup>a</sup> tissues	Cytophagales	GH35
<i>A. aquaticus</i>	Laboratory	Female	Hindgut	Bacteria	GH3
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Armatimonadetes	CE10
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH18
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH136
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH2
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH50
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	PL9_2
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	CE1
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH5_13
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH93
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH99
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Bacteria	GH135
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Phycisphaerae	GH99
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Planctomycetes	GH44
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Planctomycetes	GH26 + CBM35
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Planctomycetes	GH18
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Planctomycetes	GH43_5
<i>A. aquaticus</i>	Laboratory	Male	Hindgut	Planctomycetes	CE1
<i>A. aquaticus</i>	Field	Female	Nd. <sup>a</sup> tissues	Polynucleobacter	GT102
<i>P. dilatatus d.</i>	Laboratory	Male	Nd. <sup>a</sup> tissues	Algoriphagus	GH33
<i>P. dilatatus p.</i>	Laboratory	Female	Hindgut	Algoriphagus	CBM38
<i>P. dilatatus p.</i>	Laboratory	Female	Nd. <sup>a</sup> tissues	Bacteroidetes	CBM32
<i>P. pruinosis</i>	Field	Female	Hindgut	Bacteroidetes	CE3
<i>P. pruinosis</i>	Field	Female	Hindgut	Rhizobiales	GH108
<i>P. pruinosis</i>	Laboratory	Female	Caeca	Rhizobiales	CE10

Assignment corresponds to the closest bacterial taxon possible as predicted by CAT-BAT

<sup>a</sup> Nd. non-digestive tissues

However, its microbial CAZome comprised 43 different lignocellulose-degrading CAZymes, which is 1.5 times more than that of terrestrial species. Moreover, the freshwater isopod did not have laccases for lignin degradation, and most of the LMEs seem to be produced by its microbiota, as described in accordance with Zimmer and Bartholmé [42]. Similarly, the CAZome of its microbiota contained a significant number of hemicellulases, which could allow *A. aquaticus* to degrade the wide variety of consumed plants and fungi [43]. *Asellus aquaticus* would thus compensate its small enzymatic repertoire thanks to its microbiota which would then provide the necessary CAZymes to achieve food digestion.

We also showed that the microbiota could use cellulosomes and PULs to improve lignocellulose degradation. To our knowledge, this is the first evidence of the presence of cellulosomes and PULs in isopods. Only Zimmer [26] has suspected the presence of cellulosomes in bacteria from the hindgut of terrestrial isopods. Cellulosomes are described as “one nature’s most elaborate and highly efficient nanomachines” [44]. They are multiprotein complexes where associated enzymes collaborate to degrade cellulose and hemicellulose. Each enzyme is specifically regulated allowing for a sequential intervention, thus avoiding any competitive interactions [45]. We have identified potential cellulosomes in bacteria mainly found in the hindgut of both aquatic and terrestrial isopod hosts. Many CAZymes associated with a predicted dockerin domain are known to degrade cellulose and hemicellulose, indicating an enhanced ability for the isopod microbiota to degrade lignocellulose through cellulosomes. Identified cohesin and dockerin modules belong to several bacteria species, suggesting inter-species and intra-species cohesin–dockerin interactions [46–48]. PULs are other complexes first described in Bacteroidetes genomes for lignocellulose and other carbohydrates degradation [49]. They are organized in several co-regulated and colocalized genes encoding CAZymes, sensing proteins, binding proteins, and transporters [50]. We found several candidate contigs harboring PUL genes markers *SusC* or/and *SusD*. Among them, we have predicted several PULs involved in hemicellulose degradation and affiliated to unclassified Bacteroidetes bacteria. Nevertheless, there are probably a great number of fragmented PULs in our metagenomes due to the high proportion of *SusC* and *SusD* orphan domains identified in the contigs. Isopod holobionts are therefore good candidates for research on cellulosomes and PULs, and for the discovery of new bacteria taxa encoding these complexes.

Microbial communities linked to lignocellulose degradation are highly diversified in all host species. Most of the 43 identified bacteria families belong to Actinobacteria, Bacteroidetes and Proteobacteria phyla, where

many species are known to produce CAZymes [13]. Bacterial communities associated to lignocellulose degradation are different not only across species, but also between populations in a single species. The isopod microbiota is mostly composed of environmental bacteria and depends on several factors like environmental conditions, sex and season [22, 23, 51, 52]; intra- and inter-host diversity is thus very labile. Yet, as it has been observed from previous studies in *A. vulgare* [24, 25], there is a probable functional redundancy for the lignocellulose degradation between isopod microbiota, in particular in terrestrial host species. This functional redundancy directly reflects the holobiont concept, which considers that each biont is selected through metabolic and developmental interactions that take place in the holobiont [6]. The lignocellulose degradation is therefore an important process that shapes and drives the isopod holobiont composition, selecting the function over the individual. From this point of view, isopods differ from other well-known lignocellulose decomposers (e.g. ruminants, termites, etc.), where transient microorganisms have evolved into heritable symbionts and thus promoting the evolution of herbivory [7].

## Conclusion

As part of the holobiont concept, isopods are excellent models to study lignocellulose degradation. First, they harbor diverse and rich microbial communities in their digestive tissues, likely providing them with complementary lignocellulose-degrading CAZymes. Lignocellulose degradation is therefore possible thanks to multiple interactions between the host and its microbial bionts. Second, strategies for lignocellulose digestion vary across isopod species. In the freshwater isopod *A. aquaticus*, the contribution of the microbiota for this process is much more important than in terrestrial species. On the contrary, marine isopods of the *Limnoria* genus degrade the lignocellulose without the help of any microbiota [19]. While the latter use their hemocyanins to facilitate lignocellulose digestion, terrestrial and freshwater isopod could use PUL and cellulosome systems from their microbiota, as well as specific enzymes like LPMOs to improve their digestion of lignocellulose. It is very likely that many strategies remain to be discovered in isopods, especially in marine ones, which constitute the most abundant and yet the least studied group of species. Isopods are therefore promising models for the biotechnological industries for biomass conversion. The discovery of novel CAZymes and relevant lignocellulose-degrading bacteria strains in isopod holobionts would help promote new sustainable methods and tools to replace fossil fuels.

## Methods

### Biological samples

Metagenomic data were generated from laboratory lineages of the freshwater isopod *Asellus aquaticus*, and the terrestrial isopods *Porcellionides pruinosus*, *Porcellio dilatatus dilatatus* and *Porcellio dilatatus petiti*. Field populations of *Asellus aquaticus* from the Pinail nature reserve (France, 46° 42' 2.698" N, 0° 31' 13.378" E) and *Porcellionides pruinosus* from Nouaillé-Maupertuis (France, 46 30' 34" N, 0° 24' 54" E) were also collected in July 2017 (Additional file 1). Those individuals were kept until dissection (within 2 days after collection) in plastic boxes with water or soil from their respective sampling sites. *Armadillidium vulgare* data were issued from our previous metagenomic study [25].

### Metagenomics: DNA extraction and sequencing

Prior to dissection, all individuals were surface-sterilized using sodium hypochlorite. Tissues were then dissected out using sterilized instruments. All tissues were rinsed in Ringer solution to avoid cross-contamination between tissues. Caeca and hindguts (with their contents) were kept as separate samples, and the remaining tissues (i.e. nerve cords, gonads and fat tissues) were pooled. All samples were homogenized in extraction buffer, and total DNA was purified using phenol–chloroform [53]. Equimolar amounts of DNA from seven biological replicates (except for females of *A. aquaticus* from the Pinail nature reserve for which only six individuals could be sampled) of the same tissue and sample type (i.e. origin and sex) were pooled. Then, prokaryotic DNA was enriched twice in each pool using the NEBNext® Microbiome DNA Enrichment kit (New England Biolabs) according to the manufacturer's instructions. This resulted in 36 shotgun metagenomic libraries which were sequenced on an Illumina HiSeq 4000 by GenoScreen (Lille, France), generating 2 × 150 bp pair-end reads (Additional file 1).

### Metagenomic shotgun assembly

Read quality was checked with FastQC (version 0.11.2; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and removal of sequencing adaptors and low quality bases was performed with Trimmomatic (version 0.32; [54]). Trimmed reads shorter than 50 bp were discarded. To identify and filter rRNA reads, SortMeRNA was used with an *E* value cut-off of 1e-20 (version 2.1; [55]). To discard host reads, remaining reads were mapped using BOWTIE2 (version 2.3.4.3; [56]) against a custom isopod database comprising all isopod sequences from the Nucleotide NCBI database and unpublished isopod sequences from our laboratory. Non-mapped reads were then assembled with the MEGAHIT software (version

1.0.3; [57]) using the following parameters: –min-count 2 –k-min 21 –k-max 127 –k-steps 1. Assembly qualities were checked using Blobtools (version 1.0; [58]).

To ensure that all host sequences were filtered, two control steps were performed. (1) ORFs from contigs were predicted using Prodigal (version 2.60; [59]) with “meta” parameter, and they were compared against the Non-Redundant Protein database (December 1, 2018) using BLASTX [60] with an *E* value cut-off of 0.0001. The BLAST outputs were then imported into MEGAN6 software (version 6.15; [61]) for taxonomic assignment based on the lowest common ancestor (LCA) algorithm using the NCBI taxonomy database. All contigs associated to an ORF assigned to eukaryotes were discarded. (2) Then, the remaining contigs were run through the CAT-BAT pipeline (version v4.6; [62]) for taxonomic classification and those that were assigned to eukaryotes were discarded.

### Host transcriptomes

Host transcriptomes of whole individuals of *P. dilatatus dilatatus*, *P. dilatatus petiti*, *P. pruinosus*, *A. aquaticus* and *A. vulgare* were the same as those used in Bredon et al. [18]. In brief, reads were trimmed using Trimmomatic (version 0.32; [54]), transcriptome assemblies were performed with IDBA-TRAN [63] with default parameters, then transcripts were clustered with ≥ 95% identity using CD-HIT-EST (version 4.6; [64]) and ORFs were predicted using Transdecoder (version 3.0.1; <https://transdecoder.github.io/>). The completeness of the resulting assemblies was assessed with BUSCO (version 3.0.1; [65]) referring to core arthropod genes.

### Carbohydrate-Active enZYme annotation

CAZymes were identified using the Carbohydrate Active enZymes (CAZy) database [13]. dbCAN2 [66] was used to identify CAZy families (i.e. Glycoside Hydrolases (GHs), Glycosyl Transferases (GTs), Polysaccharide Lyases (PLs), Carbohydrate Esterases (CEs), Auxiliary Activities (AAs) and Carbohydrate-Binding Modules (CBMs)) from the previously filtered ORFs from the metagenomes and transcriptomes. The software integrates three tools for CAZymes annotation: (1) HMMER (version 3.2.1; [67]) that uses the dbCAN CAZyme domain HMM database [68] for domain predictions, (2) DIAMOND (version 0.9.24; [69]) for sequence comparisons against a custom pre-annotated CAZyme sequence database, and (3) HOTPEP [70] that performs searches against a conserved CAZyme short peptide database. dbCAN2 was run with the following parameter: –dia\_eval 1e-50. In a conservative manner, only CAZymes that were predicted by the three tools were kept for the following analyses.

As recommended by dbCAN2 authors, CAZyme counts (note that “CAZyme” refers to functional modules or domains, not genes) resulting from dbCAN assignment using HMMER tool were considered for the following analyses. For comparative analyses, CAZyme counts were normalized to even out the heterogeneity arising for differential library sizes. For each sample, CAZyme counts were divided by the number of ORFs in the metagenome or transcriptome of interest to calculate the relative abundance for each CAZyme family. Then, the normalized count of each family in each metagenome or transcriptome was calculated by multiplying the relative abundance by the lowest number of ORFs identified in corresponding datasets: 202,349 (i.e. male tissues of *P. dilatatus dilatatus*) for the newly sequenced metagenomes, 22,641 (i.e. female tissues of *A. vulgare* from the field) for *A. vulgare* metagenomes of our previous study and 19,473 (i.e. *A. aquaticus* transcriptome) for host transcriptomes.

All CAZy families known to potentially contribute to lignocellulose degradation were then selected for further analysis. However, a single CAZy family can bring together enzymes involved in a large variety of carbohydrate-modifying activities, including lignocellulose degradation. For that reason, the enzymatic activities of the CAZymes belonging to those families were predicted using Hotpep [70] in order to confirm their implication in lignocellulose degradation. When Hotpep could not predict the function of the CAZymes, we considered their most common activity reported in the CAZy database [13].

### Community profiling

To identify microbial communities encoding CAZymes, ORFs annotated as CAZymes were compared with the Non-Redundant Protein database (December 1, 2018) using BLASTP [60]. An *E* value cut-off of 0.0001 was used and the top five hits were kept. MEGAN6 software (version 6.15; [61]) was then used for taxonomic assignment of ORFs using the NCBI taxonomy database, and to construct principal component analysis (PCA) and phylogenetic tree. Results were visualized using the Phyloseq R package [71].

### PUL and cellulosome identification

PULs consist of co-localized and co-regulated genes organized around an *SusC–SusD* gene pair and encoding proteins that degrade complex carbohydrates. They might play an important role in the breakdown of lignocellulose in Bacteroidetes species [49]. To identify potential PULs in microbiota, ORFs of metagenomes were first compared to the Pfam database (version 32.0;

[72]) using hmm-search (version 3.2.1; [67]) with an *E* value cut-off of 0.0001 to identify conserved domains. Then, contigs encoding PUL gene markers [*SusD* like proteins (PF07980) and TonB-dependent receptor/*SusC* like proteins (PF00593)] were extracted and those encoding one sequential pair of *susC* and *susD* were kept for the following analyses. These sequences were subjected to CAT-BAT for taxonomic assignment, dbCAN2 for CAZyme annotation and Prokka (version 1.9; [73]) for gene annotation.

Cellulosomes are multi-enzyme lignocellulosic systems organized around a scaffolding and attached to bacterial cells [44]. Enzymes bind to the scaffolding thanks to interactions among cohesin and dockerin modules. To identify cellulosome systems in microbiota, we searched for cohesin Pfam domain (PF00963) and dockerin Pfam domain (PF00963) in the conserved domains predicted above, as well as for S-layer homology domain (SLH; PF00395). Then, contigs encoding those modules were subjected to CAT-BAT for taxonomic assignment and dbCAN2 for CAZyme annotation.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13068-020-01683-2>.

**Additional file 1.** Metrics of the metagenomic samples and resulting assemblies.

**Additional file 2.** Metrics of the transcriptomic samples and resulting transcriptomes.

**Additional file 3.** List of CAZymes identified in the metagenomes and transcriptomes assemblies.

**Additional file 4.** Protein sequences of CAZyme genes identified in the metagenome assemblies.

**Additional file 5.** Protein sequences of CAZyme genes identified in the transcriptome assemblies.

**Additional file 6.** List of predicted enzymatic functions of CAZymes identified in metagenomes and host transcriptomes.

**Additional file 7.** List of PULs identified in the metagenomes.

**Additional file 8.** List of cohesin, dockerin and SLH modules identified in metagenomes.

**Additional file 9.** Contig sequences that contain dockerin, cohesin and SLH modules.

### Abbreviations

AAs: Auxiliary activities; CAZome: CAZyme repertoire; CAZymes: Carbohydrate-active enzymes; CBMs: Carbohydrate-binding modules; CDHs: Cellobiose dehydrogenases; cDNA: Complementary deoxyribonucleic acid; CEs: Carbohydrate esterases; DNA: Deoxyribonucleic acid; GHs: Glycoside hydrolases; GTs: Glycosyltransferases; LMEs: Lignin-modifying enzymes; LPMOs: Lytic polysaccharide mono-oxygenases; NCBI: The National Center for Biotechnology Information; ORF: Open reading frame; PCA: Principal component analysis; Pfam: Protein families; PLs: Polysaccharide lyases; PULs: Polysaccharide utilization loci; RNA: Ribonucleic acid; rRNA: Ribosomal ribonucleic acid.

### Acknowledgements

We thank Carine Delaunay and Maryline Raimond for assistance with animal dissections and in RNA extractions, and Alexandra Lafitte for animal rearing. We thank Yann Sellier, Clémentine Préau and the Pinail nature reserve for their contribution to the sampling of *A. aquaticus* individuals.

### Authors' contributions

DB and MB conceived and designed the study. DB and BM supervised the study. MB performed sequence processing, data analysis and drafted the manuscript. BM supervised the bioinformatic data analysis. BH prepared RNA for sequencing and performed assemblies of the transcriptomes. JB and PG supervised the acquisition of transcriptomic data. 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 work of MB was supported by a Ph.D. 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 Additional file 1. Identified CAZymes are provided in FASTA format in Additional files 4 and 5.

### 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: 28 September 2019 Accepted: 20 February 2020

Published online: 13 March 2020

### References

- McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci*. 2013;110:3229–36.
- Blaser MJ. The microbiome revolution. *J Clin Invest*. 2014;124:4162–5.
- Margulis L, Fester R. Symbiosis as a source of evolutionary innovation: speciation and morphogenesis. Cambridge: MIT Press; 1991.
- Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. *Microbiome*. 2018. <https://doi.org/10.1186/s40168-018-0457-9>.
- Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, et al. Getting the hologenome Concept Right: an eco-evolutionary framework for hosts and their microbiomes. *mSystems*. 2016;1:00028–116.
- Doolittle WF, Booth A. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol Philos*. 2017;32:5–24.
- Gilbert SF. Developmental symbiosis facilitates the multiple origins of herbivory. *Evol Dev*. 2019. <https://doi.org/10.1111/ede.12291>.
- Ragauskas AJ. The path forward for biofuels and biomaterials. *Science*. 2006;311:484–9.
- Bomble YJ, Lin C-Y, Amore A, Wei H, Holwerda EK, Ciesielski PN, et al. Lignocellulose deconstruction in the biosphere. *Curr Opin Chem Biol*. 2017;41:61–70.
- Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA. Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels*. 2010;1:323–41.
- López-Mondéjar R, Algora C, Baldrian P. Lignocellulolytic systems of soil bacteria: a vast and diverse toolbox for biotechnological conversion processes. *Biotechnol Adv*. 2019. <https://doi.org/10.1016/j.biotechadv.2019.03.013>.
- 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.
- 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.
- Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*. 2002;66:72.
- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZY database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*. 2013;6:41.
- Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink V. Novel enzymes for the degradation of cellulose. *Biotechnol Biofuels*. 2012;5:45.
- Sabbadin F, Hemsforth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, et al. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat Commun*. 2018. <https://doi.org/10.1038/s41467-018-03142-x>.
- Bredon M, Herran B, Lheraud B, Bertaux J, Grève P, Moumen B, et al. Lignocellulose degradation in isopods: new insights into the adaptation to terrestrial life. *BMC Genomics*. 2019;20:462.
- Besser K, Malyon GP, Eborall WS, da Paro Cunha G, Filgueiras JG, Dowle A, et al. Hemocyanin facilitates lignocellulose digestion by wood-boring marine crustaceans. *Nat Commun*. 2018. <https://doi.org/10.1038/s41467-018-07575-2>.
- 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.
- Ray DL, Julian JR. Occurrence of cellulase in *Limnoria*. *Nature*. 1952;169:32–3.
- 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. <https://doi.org/10.3389/fmicb.2016.01472>.
- Dittmer J, Bouchon D. Feminizing *Wolbachia* influence microbiota composition in the terrestrial isopod *Armadillidium vulgare*. *Sci Rep*. 2018. <https://doi.org/10.1038/s41598-018-25450-4>.
- Dittmer J, Lesobre J, Moumen B, Bouchon D. Host origin and tissue microhabitat shaping the microbiota of the terrestrial isopod *Armadillidium vulgare*. *FEMS Microbiol Ecol*. 2016;92:fw063.
- Bredon M, Dittmer J, Noël C, Moumen B, Bouchon D. Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate. *Microbiome*. 2018. <https://doi.org/10.1186/s40168-018-0536-y>.
- Zimmer M. The role of animal–microbe interactions in isopod ecology and evolution. *Acta Biol Benrodis*. 2006;13:127–68.
- Vermeij GJ, Lindberg DR. Delayed herbivory and the assembly of marine benthic ecosystems. *Paleobiology*. 2000;26:419–30.
- Poore AGB, Ahlyong ST, Lowry JK, Sotka EE. Plant feeding promotes diversification in the Crustacea. *Proc Natl Acad Sci*. 2017;114:8829–34.
- Davison A, Blaxter M. Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Mol Biol Evol*. 2005;22:1273–84.
- 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.
- Terrapon N, Lombard V, Gilbert HJ, Henrissat B. Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics*. 2015;31:647–55.
- Zimmer M. Nutrition in terrestrial isopods (Isopoda: *Oniscidea*): an evolutionary-ecological approach. *Biol Rev*. 2002;77:455–93.
- Schmitz EH, Schultz TW. Digestive anatomy of terrestrial isopoda: *Armadillidium vulgare* and *Armadillidium nasatum*. *Am Midl Nat*. 1969;82:163.
- 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.
- 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.

36. 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.
37. Geng A, Cheng Y, Wang Y, Zhu D, Le Y, Wu J, et al. Transcriptome analysis of the digestive system of a wood-feeding termite (*Coptotermes formosanus*) revealed a unique mechanism for effective biomass degradation. *Biotechnol Biofuels*. 2018. <https://doi.org/10.1186/s13068-018-1015-1>.
38. Brune A. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*. 2014;12:168–80.
39. Ni J, Tokuda G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv*. 2013;31:838–50.
40. Morozova OV, Shumakovich GP, Gorbacheva MA, Shleev SV, Yaropolov AI. “Blue” laccases. *Biochem Mosc*. 2007;72:1136–50.
41. Janusz G, Pawlik A, Sulej J, Świdorska-Burek U, Jarosz-Wilkolazka A, Paszczyński A. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol Rev*. 2017;41:941–62.
42. 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.
43. 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.
44. Fontes CMGA, Gilbert HJ. Cellulosesomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem*. 2010;79:655–81.
45. Obeng EM, Adam SNN, Budiman C, Ongkudon CM, Maas R, Jose J. Lignocellulases: a review of emerging and developing enzymes, systems, and practices. *Bioresour Bioprocess*. 2017;4:16.
46. Slutzki M, Reshef D, Barak Y, Haimovitz R, Rotem-Bamberger S, Lamed R, et al. Crucial roles of single residues in binding affinity, specificity, and promiscuity in the cellulosomal cohesin–dockerin interface. *J Biol Chem*. 2015;290:13654–66.
47. Sakka K, Sugihara Y, Jindou S, Sakka M, Inagaki M, Sakka K, et al. Analysis of cohesin–dockerin interactions using mutant dockerin proteins: mutational analysis of cohesin–dockerin interactions. *FEMS Microbiol Lett*. 2011;314:75–80.
48. Pagès S, Bélaïch A, Bélaïch J-P, Morag E, Lamed R, Shoham Y, et al. Species-specificity of the cohesin–dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of specificity determinants of the dockerin domain. *Proteins Struct Funct Genet*. 1997;29:517–27.
49. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: fueling microbial communities. *J Bacteriol*. 2017. <https://doi.org/10.1128/jb.00860-16>.
50. Hemsworth GR, Dejean G, Davies GJ, Brumer H. Learning from microbial strategies for polysaccharide degradation. *Biochem Soc Trans*. 2016;44:94–108.
51. Horvathova T, Bauchinger U. Biofilm improves isopod growth independent of the dietary cellulose content. *Physiol Biochem Zool*. 2019;92:531–43.
52. Wenzel MA, Douglas A, Piertney SB. Microbiome composition within a sympatric species complex of intertidal isopods (*Jaera albifrons*). *PLoS ONE*. 2018;13:e0202212.
53. Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, et al. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci*. 1989;86:6196–200.
54. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
55. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28:3211–7.
56. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
57. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11.
58. Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Research*. 2017;6:1287.
59. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119.
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
61. Huson DH, Beier S, Flade I, Górska 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.
62. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Bioinformatics*. 2019;20:217. <https://doi.org/10.1101/530188>.
63. 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; 2010. p. 426–40. [https://doi.org/10.1007/978-3-642-12683-3\\_28](https://doi.org/10.1007/978-3-642-12683-3_28).
64. 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.
65. 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.
66. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46:W95–101.
67. 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.
68. 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.
69. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
70. 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 Bioinform*. 2017. <https://doi.org/10.1186/s12859-017-1625-9>.
71. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*. 2013;8:e61217.
72. Finn RD, Coggill 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.
73. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.

## Publisher's Note

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