Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms

Philipp Rausch\textsuperscript{1,2,3,*}, Malte Rühlemann\textsuperscript{4†}, Britt M. Hermes\textsuperscript{1,2,5}, Shaunı Doms\textsuperscript{1,2}, Tal Dagan\textsuperscript{6}, Katja Dierking\textsuperscript{7}, Hanna Domini\textsuperscript{8}, Sebastian Fraune\textsuperscript{8}, Jakob von Frieling\textsuperscript{9}, Ute Hentschel\textsuperscript{10,11}, Femke-Anouska Heinsen\textsuperscript{4}, Marc Höppner\textsuperscript{4}, Martin T. Jahn\textsuperscript{10}, Cornelia Jaspers\textsuperscript{11,12}, Kohar Annie B. Kissoyan\textsuperscript{7}, Daniela Langfeldt\textsuperscript{6}, Ateequr Rehman\textsuperscript{4}, Thorsten B. H. Reusch\textsuperscript{11,12}, Thomas Roeder\textsuperscript{9}, Ruth A. Schmitz\textsuperscript{6}, Hinrich Schulenburg\textsuperscript{7}, Ryszard Soluch\textsuperscript{6}, Felix Sommer\textsuperscript{4}, Eva Stukenbrock\textsuperscript{13,14}, Nancy Weiland-Bräuer\textsuperscript{6}, Philip Rosenstiel\textsuperscript{4}, Andre Franke\textsuperscript{4}, Thomas Bosch\textsuperscript{8} and John F. Baines\textsuperscript{1,2,*}

Abstract

Background: The interplay between hosts and their associated microbiome is now recognized as a fundamental basis of the ecology, evolution, and development of both players. These interdependencies inspired a new view of multicellular organisms as “metaorganisms.” The goal of the Collaborative Research Center “Origin and Function of Metaorganisms” is to understand why and how microbial communities form long-term associations with hosts from diverse taxonomic groups, ranging from sponges to humans in addition to plants.

Methods: In order to optimize the choice of analysis procedures, which may differ according to the host organism and question at hand, we systematically compared the two main technical approaches for profiling microbial communities, 16S rRNA gene amplicon and metagenomic shotgun sequencing across our panel of ten host taxa. This includes two commonly used 16S rRNA gene regions and two amplification procedures, thus totaling five different microbial profiles per host sample.

Conclusion: While 16S rRNA gene-based analyses are subject to much skepticism, we demonstrate that many aspects of bacterial community characterization are consistent across methods. The resulting insight facilitates the selection of appropriate methods across a wide range of host taxa. Overall, we recommend single- over multi-step amplification procedures, and although exceptions and trade-offs exist, the V3 V4 over the V1 V2 region of the 16S rRNA gene. Finally, by contrasting taxonomic and functional profiles and performing phylogenetic analysis, we provide important and novel insight into broad evolutionary patterns among metaorganisms, whereby the transition of animals from an aquatic to a terrestrial habitat marks a major event in the evolution of host-associated microbial composition.

Keywords: Animal microbiome, Evolution, Phylosymbiosis, Holobiont, Metaorganism
Background
Dynamic host-microbe interactions have shaped the evolution of life. Virtually all plants and animals are colonized by an interdependent complex of microorganisms, and there is growing recognition that the biological processes of hosts and their associated microbial communities function in tandem, often as biological partners comprising a collective entity known as the metaorganism [1]. For instance, symbiotic bacteria contribute to host health and development in critical ways, ranging from nutrient metabolism to regulating whole life cycles [2] and in turn benefit from habitats and resources the host provides. Moreover, it is well established that perturbations of the microbiome likely play an important role in many host disease states [3]. However, researchers have yet to elucidate the mechanisms driving these interactions, as the exact molecular and cellular processes are only poorly understood.

An integrated view on the metaorganism encompasses a cross-disciplinary approach that addresses how and why microbial communities form long-term associations with their hosts. Despite widespread agreement that the interdependencies of microbes and their hosts warrant study, there remains considerable incongruity between researchers regarding the best methodologies to study host-microbe interactions. The development of standardized protocols for characterizing and analyzing host-associated microbiomes across the tree of life is thus crucial to understand the evolution and function of metaorganisms without the issues of technical inconsistencies or data quality.

The rapidly growing interest in microbiome research has been bolstered by the ability to profile diverse microbial communities using next-generation sequencing (NGS). This culture-free, high-throughput technology enables identification and comparison of entire microbial communities, so-called metagenomics [4]. Metagenomics typically encompasses two particular sequencing strategies: amplicon sequencing, most often of the 16S rRNA gene as a phylogenetic marker; or shotgun sequencing, which captures the complete breadth of DNA within a sample [4].

The use of the 16S ribosomal RNA gene as a phylogenetic marker has proven to be an efficient and cost-effective strategy for microbiome analysis and even allows for the imputation of functional content based on taxon abundances [5]. However, PCR-based phylogenetic marker protocols are vulnerable to biases through sample preparation and sequencing errors. The choice of which hypervariable regions of the 16S rRNA gene are targeted for sequencing seems to be among the biggest factors underlying technical differences in microbiome composition [6–8]. Furthermore, 16S rRNA gene amplicon sequencing is typically limited to taxonomic classification at the genus level depending on the database and classifiers used [9], and provides only limited functional information [5]. These well-recognized limitations of amplicon-based microbial community analyses have raised concerns about the accuracy and reproducibility of 16S rRNA phylogenetic marker studies and have led to an increased interest in developing more reliable methods for amplicon library preparation and sequencing [8, 10].

 Shotgun metagenomics, on the other hand, offers the advantage of species- and strain-level classification of bacteria. Additionally, it allows researchers to examine the functional relationships between hosts and bacteria by determining the functional content of samples directly [9, 11], and enables the exploration of yet unknown microbial life that would otherwise remain unclassifiable [12]. However, the relatively high costs of shotgun metagenomics and more demanding bioinformatic requirements have precluded its use for microbiome analysis on a wide scale [4, 9].

In this study, we set out to systematically compare experimental and analytical aspects of the two main technical approaches for microbial communities profiling. 16S rRNA gene amplicon and shotgun sequencing, across a diverse array of host species studied in the Collaborative Research Center 1182, “Origin and Function of Metaorganisms.” The ten host species range from basal aquatic metazoans [Aplysina aerophoba (sponge) and Mnemiopsis leidyi (comb jelly)]; to marine and limnic cnidarians [Aurelia aurita, Nematostella vectensis, Hydra vulgaris], standard vertebrate (Mus musculus), and invertebrate model organisms (Drosophila melanogaster, Caenorhabditis elegans); to Homo sapiens; and in addition to wheat (Triticum aestivum) and a standardized mock community. This setup provides a breadth of samples in terms of taxonomic composition and diversity. Conducting standardized data generation procedures on these diverse samples on the one hand provides a unique and powerful opportunity to systematically compare alternative methods, which display considerable heterogeneity in performance. On the other hand, this information enables researchers working on these or similar host species to choose the experimental (e.g., hypervariable region) or analytical pipelines that best suit their needs, which will be a valuable resource to the greater community of host-microbe researchers. Finally, we identified a number of interesting, broad-scale patterns contrasting the aquatic and terrestrial environment of metaorganisms, which also reflect their evolutionary trajectories.

Results
Our panel of hosts includes ten species, for which five biological replicates each were included (see
Additional file 1: Figure S1). The majority of hosts are metazoans, including the “golden sponge” (*Aplysina aerophoba*), moon jellyfish (*Aurelia aurita*), comb jellyfish (*Mnemiopsis leidyi*), starlet sea anemone (*Nematostella vectensis*), fresh-water polyp *Hydra vulgaris*, roundworm (*Ceanorhabditis elegans*), fruit fly (*Drosophila melanogaster*), mouse (*Mus musculus*), human (* Homo sapiens*), and the inclusion of wheat (*Triticum aestivum*), which can serve as an outgroup to the metazoan taxa. *Drosophila melanogaster* was additionally sampled using two different methods targeting feces and intestinal tissue. Nucleic acid extraction procedures were conducted according to the needs of the individual host species (see the “Methods” section and Additional file 1), after which all DNA templates were subjected to a standard panel of sequencing procedures. For 16S rRNA gene amplicon sequencing, we used primers flanking two commonly used variable regions, the V1 V2 and V3 V4 regions. Further, for each region, we compared a single-step fusion-primer PCR to a two-step procedure designed to improve the accuracy of amplicon-based studies [8]. Finally, all samples were also subjected to shotgun sequencing, such that five different sequence profiles were generated for each sample. While a single classification pipeline was employed for all four 16S rRNA gene amplicon sequence profiles, community composition based on shotgun data was evaluated using MEGAN [13], due to the advantage of simultaneously performing taxonomical and functional classification of shotgun-based techniques, we plotted the discovered classifiable taxa and functions for the entire pooled dataset. Although the methods differ distinctly, each method shows a plateau in the number of discovered entities (see Additional file 1: Figures S3C, S3D).

**Mock community**

The analysis of standardized mock communities is an important measure to ensure general quality standards in microbial community analysis. In this study, we employed a commercially available mixture of eight bacterial and two yeast species. Comparison among the amplification procedures (one- and two-step PCR), 16S rRNA gene regions (V1 V2, V3 V4), and shotgun data reveals varying degrees of similarity to the expected microbial community composition (Fig. 1). One discrepancy is apparent due to the misclassification of *Escherichia/Shigella*, whose close relationship makes delineation at the genus level difficult based on the V1 V2 region and is subsequently classified to *Enterobacteriaceae* (Fig. 1a, Additional file 1: Figure S4). Classification of this bacterial group also differs based on the shotgun analysis employed, due to different naming and taxonomic standards of the respective databases (*Escherichia, Shigella*, and *Enterobacteriaceae* refer to the *Escherichia/Shigella* cluster) [14]. However, overall, the amplicon-based profiles show the closest matches to the expected community. The V3 V4 one-step method shows the lowest degree of deviation between observed and expected abundances of the focus taxa (Table 1; Additional file 1: Figure S4). In addition, the relative abundances of fungi in the mock community were relatively well predicted by MEGAN (see Fig. 1).

Next, we evaluated alpha and beta diversity across the different technical and analytical methods. Interestingly, most methods overestimate taxon richness but
underestimate complexity (as measured by the Shannon index) of the mock community, which could reflect biases arising from grouping taxon abundances based on slightly differing taxonomies (Fig. 1c, Additional file 1: Figures S4, S5A and Additional file 2: Table S2). Overall, the amplicon methods appear to more accurately reflect alpha diversity, although significant differences are present with regard to the amplified region (species richness: variable region $F_{1,10} = 6.3657, P = 0.0302$; Shannon H: method $F_{1,9} = 3.330, P = 0.1014$, variable region $F_{1,9} = 6.110, P = 0.0354$; ANOVA best model). With regard to beta diversity, the largest distance to the expected composition is observed for the shotgun-based data, while the amplicon-based techniques, in particular V3 V4, show the lowest distance (Fig. 1d, Additional file 1: Figure S5B). Pairwise tests show almost no differences between the amplicon-based techniques, while the shotgun-based data significantly differs from all amplicon profiles (Additional file 2: Table S3). Thus, in conclusion, shotgun-based analysis yields a higher degree of error compared to the amplicon-based approaches for the simple mock community used in our study.

**Taxonomic diversity within and between hosts**

To evaluate the performance of our panel of metagenomic methods over the range of complex host-associated communities in our consortium, we next employed a series of alpha and beta diversity analyses to
these samples, which also provides an opportunity to infer broad patterns across animal taxa based on a standardized methodology. Measures of alpha diversity display overall consistent values with respect to host species, although many significant differences between methods are present, which are mostly host-specific (Fig. 2a, b). However, several host taxa display high levels of consistency across methods including *A. aurita*, *C. elegans*, *D. melanogaster*, and *H. sapiens*, which show almost no significant differences between methods. Discrepancies and individual recommendations for each host species are discussed in Additional file 1: Figures S6–S16 and Additional file 2: Table S4. An intriguing observation is the tendency of aquatic hosts to display higher alpha diversity values than those of terrestrial hosts, which is supported by average differences between aquatic and terrestrial hosts and by relative consistent comparisons among single host species as well (Fig. 2c, d; Additional file 2: Table S5).

In order to investigate broad patterns of bacterial community similarity according to metagenomic procedure and host species, we performed beta diversity analyses including all host samples and each of their five different methodological profiles. This analysis reveals an overall strong signal of host species, irrespective of the method used to generate community profiles (Table 2; Fig. 3). Pairwise comparisons between hosts are significant in all cases except for samples derived from the V3 V4 two-step protocol, which did not consistently reach significance after correction for multiple testing (Additional file 2: Table S6). Further, complementary to the observations made for alpha diversity, we also find strong signals of community differentiation between the aquatic and terrestrial hosts (Table 2; Fig. 3b, d). The separation between these environments appears to be stronger based on amplicon data, whereas the separation between hosts is stronger based on shotgun-derived data (Table 2). To further evaluate the variability among biological replicates, we evaluated intra-group distances according to host species, which reveals organisms with generally higher community variability (i.e., *C. elegans*, *A. aurita*, *H. sapiens*, *T. aestivum*, and *M. leidyi*) than other host organisms in our study (*N. vectensis*, *M. musculus*, *D. melanogaster*, and *A. aerophoba*; Additional file 1: Figure S17A and C). Interestingly, intra-group distances also significantly differ between the aquatic and terrestrial environments, whereby aquatic organisms tend to display less variable communities than terrestrial ones (Additional file 1: Figure S17B and D). Thus, this suggests higher sample sizes may be necessary for experimental analysis of the higher variability/terrestrial taxa. The low performance of *T. aestivum* in subsequent analyses possibly originates from its commercial origin and low bacterial biomass relative to host material.

To identify individual drivers behind patterns of beta diversity, we performed indicator species analysis [15] at the genus level with respect to method, host species, and environment. Based on the amplicon data, we identified 56 of 313 indicators to display consistent associations across all four amplicon techniques, such as *Bacteroides*, *Barnesiella*, *Clostridium IV*, and *Faecalibacterium* in *H. sapiens* and *Helicobacter* and *Mucispirillum* in *M. musculus*, whereas other associations were limited to, e.g., only one variable region (Additional file 2: Tables S7 and S8). However, the overall pattern of host associations is largely consistent across methods (Additional file 1: Figure S18). We also identified numerous indicator genera for aquatic and terrestrial hosts (Additional file 2: Tables S9 and S10). Indicator analyses based on shotgun data reveals a smaller and less diverse set of host-specific indicators, which however show many congruencies with the amplicon-based data.

| Members mock community | Shot gun MEGAN | Ampicon V1 V2 one-step | V3 V4 one-step | V1 V2 two-step | V3 V4 two-step |
|-------------------------|---------------|-------------------------|----------------|---------------|---------------|
| Staphylococcus           | 0.00002       | 0.00210                 | 0.03994        | 0.21564       |
| *Listeria*               | 0.00139       | 0.00001                 | 0.00001        | 0.00001       |
| *Bacillus*               | 0.00006       | 0.00120                 | 0.00120        | 0.00120       |
| *Pseudomonas*            | 0.13668       | 0.36721                 | 0.36721        | 0.36721       |
| *Escherichia/Shigella*   | NA            | 0.00462                 | 0.00120        | 0.00120       |
| *Shigella*               | 4.6372 × 10−10| NA                      | NA             | NA            |
| *Escherichia*            | 0.00001       | NA                      | NA             | NA            |
| *Enterobacteriaceae*     | NA            | 0.87898                 | 0.87898        | 0.87898       |
| *Salmonella*             | 3.8092 × 10−5 | 0.34964                 | 0.34964        | 0.34964       |
| *Lactobacillus*          | 0.00297       | 0.87898                 | 0.87898        | 0.87898       |
| *Enterococcus*           | 0.00012       | 0.00481                 | 0.00481        | 0.00481       |

*aEscherichia/Shigella relatives counted as equivalent*
Functional diversity within and between hosts

To examine the diversity (gene richness) of metagenomic functions across host species, we evaluated EggNOG annotations (evolutionary genealogy of genes: Non-supervised Orthologous Groups [16]) to obtain a general functional spectrum (assembly-based and MEGAN), in addition to annotations derived from a database dedicated to functions interacting with carbohydrates (CAZY—Carbohydrate-Active enZymes) [17]. Overall, the individual host communities differ...
drastically in gene richness (EggNOG genes (MEGAN) \( \chi^2 = 52.202, P < 2.10 \times 10^{-16} \); EggNOG genes (assembly) \( \chi^2 = 49.986, P < 2.10 \times 10^{-16} \); CAZY \( \chi^2 = 48.815, P < 2.10 \times 10^{-16} \); approximate Kruskal-Wallis test). Although the values also differ considerably between methods, overall, the functional repertoires are most diverse in the vertebrate hosts, while only \( H. vulgaris \) and \( A. aerophoba \) as aquatic hosts carry comparably diverse functional repertoires (Fig. 4a, : Figure S19). Interestingly, in contrast to taxonomic diversity, we observe no difference in functional diversity between aquatic and terrestrial hosts.

Next we examined community differences (beta diversity) at the functional level, which are overall more pronounced (average adj. \( R^2 = 0.5084 \); Fig. 4) than those based on taxonomic (genus level) classification (shotgun adj. \( R^2 = 0.4756 \), amplicon average adj. \( R^2 = 0.4594 \); see Tables 2 and 3; Figs. 3 and 4, Additional file 1: Figure S20). On the functional level, aquatic and terrestrial hosts are considerably less distinct than observed at the taxonomic level (taxonomic shotgun data \( R^2 = 0.0766 \), taxonomic amplicon average adj. \( R^2 = 0.0690 \), functional shotgun data \( R^2 = 0.0441 \); see Tables 2 and 3; Fig. 4, Additional file 1: Figure S20). Variability of the functional repertoires was lowest in \( A. aerophoba \), \( D. melanogaster \) feces, and \( M. musculus \) gut contents, while \( H. vulgaris \), \( C. elegans \), and \( D. melanogaster \) gut samples displayed the highest intra-group distances, which translates to a higher amount of functional heterogeneity between replicates (Additional file 1: Figure S21). This reflects in large part the patterns we observed in taxonomic variability of those host-associated communities (Additional file 1: Figure S17).

### Table 2: Taxonomic distance-based PERMANOVA results for differences in community composition (genus level) between host species and host environments based on shared abundance (Bray-Curtis) and shared presence (Jaccard), based on whole genome shotgun and different amplicon strategies (\( P \) values are adjusted via Hommel’s procedure)

| Distance Factor | Data Classifier | Data | Classifier | \( DF \) | \( F \) | \( P \) | \( P_{\text{nominal}} \) | \( R^2 \) | \( \text{adj.} \ R^2 \) |
|-----------------|----------------|------|------------|---------|---------|-------|----------------|-------|-------------|
| Bray-Curtis     | Organism Shotgun | MEGAN | 10,49 | 6.3517 | 0.0001 | 0.0001 | 0.5645 | 0.4756 |
| Amplicon       | V1 V2 one-step | 10,43 | 7.1026 | 0.0001 | 0.0001 | 0.6229 | 0.5352 |
| Amplicon       | V1 V2 two-step | 10,42 | 4.2297 | 0.0001 | 0.0001 | 0.5018 | 0.3831 |
| Amplicon       | V3 V4 one-step | 10,43 | 7.8964 | 0.0001 | 0.0001 | 0.6474 | 0.5654 |
| Amplicon       | V3 V4 two-step | 10,41 | 3.7917 | 0.0001 | 0.0001 | 0.4805 | 0.3538 |
| Environment    | Organism Shotgun | MEGAN | 1,58  | 5.8958 | 0.0001 | 0.0004 | 0.0923 | 0.0766 |
| Amplicon       | V1 V2 one-step | 1,52  | 6.1588 | 0.0001 | 0.0001 | 0.1059 | 0.0887 |
| Amplicon       | V1 V2 two-step | 1,51  | 4.6185 | 0.0001 | 0.0001 | 0.0830 | 0.0651 |
| Amplicon       | V3 V4 one-step | 1,52  | 5.4975 | 0.0001 | 0.0001 | 0.0956 | 0.0782 |
| Amplicon       | V3 V4 two-step | 1,50  | 3.3349 | 0.0001 | 0.0001 | 0.0625 | 0.0438 |
| Jaccard        | Organism Shotgun | MEGAN | 10,49 | 4.7438 | 0.0001 | 0.0001 | 0.4920 | 0.3883 |
| Amplicon       | V1 V2 one-step | 10,43 | 3.6867 | 0.0001 | 0.0001 | 0.4616 | 0.3364 |
| Amplicon       | V1 V2 two-step | 10,42 | 2.9760 | 0.0001 | 0.0001 | 0.4147 | 0.2754 |
| Amplicon       | V3 V4 one-step | 10,43 | 4.0248 | 0.0001 | 0.0001 | 0.4835 | 0.3633 |
| Amplicon       | V3 V4 two-step | 10,41 | 2.9343 | 0.0001 | 0.0001 | 0.4171 | 0.2750 |
| Environment    | Organism Shotgun | MEGAN | 1,58  | 4.3872 | 0.0001 | 0.0004 | 0.0703 | 0.0543 |
| Amplicon       | V1 V2 one-step | 1,52  | 3.8714 | 0.0001 | 0.0001 | 0.0693 | 0.0514 |
| Amplicon       | V1 V2 two-step | 1,51  | 3.6541 | 0.0001 | 0.0001 | 0.0669 | 0.0486 |
| Amplicon       | V3 V4 one-step | 1,52  | 4.3213 | 0.0001 | 0.0001 | 0.0767 | 0.0590 |
| Amplicon       | V3 V4 two-step | 1,50  | 3.6646 | 0.0001 | 0.0001 | 0.0683 | 0.0497 |

### Indicator functions
To identify specific functions that are characteristic of individual hosts, we applied indicator analysis to genomic functions. General functions in EggNOG reveal several interesting patterns, including CRISPR-related genes in \( A. aerophoba \), \( H. sapiens \), and \( H. vulgaris \), suggesting a particular importance of viruses in these communities. Further, most species show characteristic genes mainly involved in energy production and conversion, amino acid transport and metabolism, replication, recombination, and repair, as well as cell wall/membrane/envelope biogenesis (Additional file 2: Tables S11–S13).

Analysis of carbohydrate-metabolizing functions based on CAZY[17] (Carbohydrate-Active enZymes) reveals the highest number of characteristic glycoside hydrolases (GH) in \( H. sapiens \) and \( M. musculus \), whereas polysaccharide lyases (PLs) for non-hydrolytic cleavage of...
glycosidic bonds are present in *A. aerophoba* and *H. sapiens* (Additional file 2: Table S14). Interestingly, parts of the cellulosome are only associated to *A. aerophoba*, while the freshwater polyp *H. vulgaris* carries characteristic auxiliary CAZYs involved specifically in lignin and chitin digestion, which may reflect adaptations of the host microbial communities to their diets (e.g., *Artemia nauplii*).

**Fig. 3** Non-metric multidimensional scaling of Bray-Curtis distances based on genus abundance profiles derived from the different 16S rRNA gene amplicon methods (V1/V2/V3/V4, one-step/two-step) and shotgun-derived genus profiles highlighting a host differences and b differences between host environments (terrestrial/aquatic; see Table 2). Non-metric multidimensional scaling of Jaccard distances based on genus presence/absence profiles derived from the different 16S rRNA gene amplicon methods and shotgun-derived genus profiles highlighting c host taxon differences and d differences between host environments (terrestrial/aquatic; see Table 2). Both panels show a separation based on host organisms and environments and not by method. Large symbols indicate the centroid of the respective host groups and vertical lines help to determine their position in space. Sample sizes are equal to Fig. 2 (see also Additional file 2: Table S1).

**Performance of metagenome imputation from 16S rRNA gene amplicon data using PICRUSt across metaorganisms**

Researchers often desire to obtain the insight gained from functional metagenomic information despite being limited to 16S rRNA gene data, for which imputation methods such as PICRUSt can be employed [5]. However, due to their dependence on variable region and database coverage [5], these imputations should be
viewed with caution. Given our dataset of both 16S amplicon and shotgun metagenomic sequences, we systematically evaluated the performance of PICRUST predictions across hosts and amplicon data type (V1 V2/V3 V4, one-step/two-step protocol). Beginning with the mock community, the V1 V2 region displays lower performance for imputing functions compared to V3 V4, as indicated by a higher weighted Nearest Sequenced Taxon Index (NSTI) \( (t = 17.812, P = 1.119 \times 10^{-7}; \) Additional file 1: Figure S22A). High NSTI values imply low availability of genome representatives for the respective sample, due to either large phylogenetic distance for each OTU to its closest sequenced reference genome or a high frequency of poorly represented OTUs [5]. Comparing the distribution of functional categories based on Clusters of Orthologous Groups (COG) [18] between

---

**Fig. 4** Functional diversities were derived from the number and abundances of MEGAN-based EggNOG annotations. Functional richness between a host organisms and b host environmental groups based is displayed, as well as functional differences between hosts (c) and environmental groups (d). Non-metric multidimensional scaling is based on Bray-Curtis distances on the differences in functional composition between the host organisms displayed (c, d, see Table 3). Large symbols indicate the centroid of the respective groups. Functional variation of communities based on pairwise Bray-Curtis distances within host organism groups and environmental groups. Sample sizes for the host taxa are \( N = 5 \), except for D. melanogaster gut tissue (\( N = 10 \); see Additional file 2: Table S1).
Table 3 Functional distance-based PERMANOVA results for differences in general functional community composition (EggNOG) and carbohydrate-active enzymes (CAZY) between host species and host environments based on shared abundance (Bray-Curtis) and shared presence (Jaccard) of functions (P values are adjusted via Hommel procedure)

| Distance Factor | Data          | DF   | F   | P       | P_{Hommel} | $R^2$ | adj. $R^2$ |
|-----------------|---------------|------|-----|---------|------------|-------|------------|
| Bray-Curtis     | Organism CAZY| 10,47 | 7.3323 | 0.0001 | 0.0001 | 0.6094 | 0.5263 |
|                 | EggNOG categories | 10,49 | 5.6088 | 0.0001 | 0.0001 | 0.5337 | 0.4386 |
|                 | EggNOG gene + description | 10,49 | 4.4454 | 0.0001 | 0.0001 | 0.4757 | 0.3687 |
|                 | EggNOG (MEGAN categories) | 10,49 | 12.2594 | 0.0001 | 0.0001 | 0.7144 | 0.6562 |
|                 | EggNOG (MEGAN gene) | 10,49 | 8.2788 | 0.0001 | 0.0001 | 0.6282 | 0.5523 |
| Environment     | CAZY         | 1,56  | 5.4257 | 0.0001 | 0.0001 | 0.0883 | 0.0721 |
|                 | EggNOG categories | 1,58  | 2.5429 | 0.0195 | 0.0195 | 0.0420 | 0.0255 |
|                 | EggNOG gene + description | 1,58  | 3.0662 | 0.0001 | 0.0007 | 0.0502 | 0.0338 |
|                 | EggNOG (MEGAN categories) | 1,58  | 3.7703 | 0.0015 | 0.0030 | 0.0610 | 0.0448 |
|                 | EggNOG (MEGAN gene) | 1,58  | 3.7271 | 0.0002 | 0.0012 | 0.0604 | 0.0442 |
| Jaccard         | Organism CAZY| 10,47 | 3.9098 | 0.0001 | 0.0001 | 0.4541 | 0.3380 |
|                 | EggNOG categories | 10,49 | 3.7179 | 0.0001 | 0.0001 | 0.4314 | 0.3154 |
|                 | EggNOG gene + description | 10,49 | 2.5275 | 0.0001 | 0.0001 | 0.3403 | 0.2057 |
|                 | EggNOG (MEGAN categories) | 10,49 | 7.7781 | 0.0001 | 0.0001 | 0.6135 | 0.5346 |
|                 | EggNOG (MEGAN gene) | 10,49 | 5.4989 | 0.0001 | 0.0001 | 0.5288 | 0.4326 |
| Environment     | CAZY         | 1,56  | 2.5866 | 0.0003 | 0.0021 | 0.0442 | 0.0271 |
|                 | EggNOG categories | 1,58  | 1.4180 | 0.1442 | 0.1442 | 0.0239 | 0.0070 |
|                 | EggNOG gene + description | 1,58  | 1.9535 | 0.0004 | 0.0024 | 0.0326 | 0.0159 |
|                 | EggNOG (MEGAN categories) | 1,58  | 3.0425 | 0.0460 | 0.0920 | 0.0498 | 0.0335 |
|                 | EggNOG (MEGAN gene) | 1,58  | 3.1222 | 0.0001 | 0.0009 | 0.0511 | 0.0347 |

Fig. 5 a Differences in the Nearest Sequenced Taxon Index (imputation success) between variable regions (average: $Z = 0.3869$, $P = 0.7017$, approximate Wilcoxon test; probability: odds ratio = 1.5941, $P = 0.1402$, Fisher’s test) and amplification method ($Z = 0.0667$, $P = 0.9472$, approximate Wilcoxon test; probability: odds ratio = 1.5511, $P = 0.1436$, Fisher’s test). b Procrustes correlation of imputed- and shotgun-based COG categories among different techniques, with significantly higher correspondence between imputed and measured functional profiles in the V1 V2 compared to the V3 V4 region ($F_{1,18} = 7.8537$, $P = 0.0118$, ANOVA). c Non-metric multidimensional scaling displays Bray-Curtis distances based on functional category abundances (COG categories) derived from PICRUSt (V1 V2/V3 V4, one-step/two-step) and shotgun-based approaches (MEGAN, single assembly). Ellipses represent standard deviations of points within the respective groups.
the different imputations (no cutoff applied) and the actual shotgun-based repertoires reveals considerable overlap except categories R (general function prediction only) and S (function unknown) (Additional file 1: Figure S22B).

Next we evaluated functional imputations for the different host species and amplification methods. We found no significant difference in average NSTI values or prediction success (NSTI < 0.15) between amplification protocols or variable region. However, approximately a third (31.8%) of the samples are lost due to incomplete imputation (NSTI > 0.15; Fig. 5a). Notable problematic host taxa are A. aerophoba and H. vulgaris, for which no sample remained below the NSTI cutoff value. Other host taxa displayed clear differential performance with regard to the variable region used, whereby H. sapiens, N. vectensis, and T. aestivum were successfully predicted based on V3 V4, but not V1 V2. However, when we employ Procrustes tests to compare community functional profiles based on shotgun sequencing (single assembly, MEGAN) and functional imputations at the COG-category level, we find a lower correspondence of the V3 V4-based imputations compared to those based on V1 V2 (Fig. 5b), while the amplification methods displayed no significant difference. A similar pattern is observed when we correlate community differences based on shotgun results and lower level (single functions) COG annotations based on PICRUSt, although the difference is not significant ($F_{1,18} = 0.6172, P = 0.4423$; ANOVA).

To investigate the similarities among methods in more detail, we merged shotgun and PICRUSt-based annotations at the level of COG categories. Principle coordinate analysis reveals only small differences between imputations with regard to amplification method or variable region (Fig. 5c). However, large differences exist between the PICRUSt and shotgun-based functional repertoires, as well as between the shotgun techniques (MEGAN, single assembly). Differences between the shotgun techniques were significant but smaller than their distance to the imputed functional spectra (Fig. 5c; Additional file 2: Table S15), a pattern also found in the relative abundances of functional categories (Additional file 1: Figure S23).

In summary, the PICRUSt-imputed functional repertoires significantly differ from actual shotgun profiles. While variation in imputation success is largely dependent on the composition of the particular host community, V3 V4 appears to more often yield successful imputations. However, when successful, V1 V2-derived imputations display closer similarity to actual functional profiles. Finally, the amplification method (one-step, two-step) appears to have no significant effect on the quality of functional imputations. These data therefore support the notion that metagenome imputations should be evaluated with care, as they depend on the underlying variable region and sample source.

**Phylogenetic patterns in microbial community composition**

The term “phylosymbiosis” refers to the phenomenon where the pattern of similarity among host-associated microbial communities parallels the phylogeny of their hosts [19]. Highly divergent hosts with drastic differences in physiology and life history might be expected to overwhelm the likelihood of observing phylosymbiosis, which can typically be observed within a given host clade [19]. However, the factors driving differences in composition among our panel of hosts may also be expected to vary in terms of the bacterial phylogenetic scale at which they are most readily observed [20]. Thus, we evaluated the degree to which bacterial community relationships (beta diversity) reflect the underlying phylogeny of our hosts at a range of bacterial taxonomic ranks, spanning from the genus to the phylum level.

In order to assess the general overlap between beta diversity and phylogenetic distance of the host species, we performed Procrustes analysis [21]. These analyses reveal that the strongest phylogenetic signal is observed when bacterial taxa are grouped at the order and/or family level, whereby the one-step protocols and the V3 V4 region display greater correlations to phylogenetic distance (Fig. 6). A similar pattern is observed for shotgun-based community profiles (i.e., MEGAN), although its fit increases again at the genus level. Measuring beta diversity based on co-occurrence of bacterial taxa between hosts (Jaccard; Fig. 6a) displays a weaker correspondence to host phylogeny than the abundance-based measure (Bray-Curtis; Fig. 6b).

To assess the fit of individual host taxa, we examined the residuals of the correlation between community composition and phylogenetic distance. This reveals a large variation in correspondence among host taxa, with M. musculus, M. leidy, H. sapiens, and D. melanogaster (feces) displaying the highest, while H. vulgaris, C. elegans, and A. aerophoba display the lowest correspondence between their microbiome composition and phylogenetic position (largest residuals; Additional file 1: Figure S24), pointing towards increased environmental influences on these microbial communities. Furthermore, terrestrial hosts display an overall better correspondence between co-occurrences of bacterial genera and host relatedness (V1 V2 one-step: $Z = 2.9578, P = 0.0025$), as do measurements based on V3 V4 (one-step: $Z = 2.7496, P = 0.0054$; two-step: $Z = 2.8097, P = 0.0046$; approximate Wilcoxon test).

Next, given the peak of correspondence between bacterial community composition and host phylogeny
observed at the order and/or family level, we set out to identify individual community members whose abundances best correlate to host phylogenetic distance using Moran’s I eigenvector method [22]. This reveals 41 bacterial families and 36 orders with significant phylogenetic signal based on one or more amplicon data set, whereby 16 families and 18 orders display repeated associations across methods (e.g., Clostridia, Bacteroidales, Desulfovibrionales; Additional file 2: Table S16; Additional file 1: Figures S25 and S26). Analyzing communities based on shotgun data on the other hand identifies 75 bacterial families and 19 orders associated with phylogenetic distances, whereby 17 and 20 display repeated associations, respectively (Additional file 2: Table S16; Additional file 1: Figure S27). The combined results of these analyses identify several families and orders with strong and consistent phylogenetic associations, in particular for the vertebrate hosts (e.g., Bacteroidaceae/Bacteroidales, Bifidobacteriaceae/Bifidobacteriales, Desulfovibrionaceae/Desulfovibrionales, Ruminococcaceae/Clostridiales; see Additional file 2: Table S16). Other individual examples include bacteria related to Helicobacteraceae/Campylobacterales in A. aurita, which are observed in other marine cnidarians and may be involved in sulfur oxidation [23]. Alcanivoracaceae, an alkane-degrading bacterial group, is strongly associated to the coastal cnidarian N. vectensis. This association might originate from adaptation to a polluted coastal environment [24]. Acidobacteria Gp6 and Gp9 specifically occur in A. aerophoba and are commonly associated to the core microbial community of sponges [25].

**Phylogenetic patterns in functional community composition**

In order to contrast the patterns observed at the taxonomic level to those based on function, we used Procrustes correlation to measure the overlap between phylogenetic distance and community distance based on the panel of functional categories in our analyses. Interestingly, the two functional categories displaying the greatest correspondence to host phylogeny are the CAZY and single EggNOG-based functions (Fig. 6). The remainder of patterns between phylogeny and bacterial functional spectra differed among the host species and functional categories (Additional file 1: Figure S28), and T. aestivum and D. melanogaster (feces) display the lowest correspondence, while C. elegans, M. musculus, and H. sapiens display the best correspondence (smallest residuals; Additional file 1: Figure S24) between their functional repertoire and phylogenetic position. As observed for the taxonomic analyses, terrestrial hosts again display a slightly better correlation than aquatic hosts (smaller residuals), in particular for the co-abundance of
EggNOG categories ($Z = 2.2116, P = 0.0267$), CAZY ($Z = 2.0393, P = 0.0414$), and the co-occurrence of EggNOG categories ($Z = 2.7377, P = 0.0061$) and genes ($Z = 3.3062, P = 0.0007$; approximate Wilcoxon test) among hosts.

Finally, to reveal individual functions correlating to host phylogeny, we used the aforementioned Moran’s $I$ eigenvector analyses with additional indicator analyses to narrow the potential clade associations. Interestingly, most functions that correlate to a specific host taxon/clade (1–3 host taxa) are mainly restricted to vertebrate hosts or in combination with a vertebrate host (Additional file 2: Tables S17–S20). This pattern is repeated across all functional annotations used in this study. Examples include fucosyltransferases, fucosidases, and polysaccharide-binding proteins, as well as different lyases for hyaluronate, xanthan, and chondroitin that stem from CAZY (see Additional file 1: Figure S28; Additional file 2: Table S17). These functions are related to glycan and mucin degradation and interaction, which mediate many intimate host-bacterial interactions and are also observed in subsequent analyses based on general functional databases (EggNOG; Additional file 2: Tables S18–S20). Many other phylogenetically correlated functions appear to be driven by the vertebrate hosts as well, which likely reflects the high functional diversity within this group (Fig. 4 and Additional file 1: Figure S21). Only LPXC and LPXK (EggNOG), genes involved in the biosynthesis of the outer membrane, are exclusively associated to the non-vertebrate hosts (LPXC, UDP-3-O-acetyl-N-acetylgalactosamine deacetylase; LPXK, Tetraacylcyldisaccharide 4′-kinase), as is an oxidative damage repair function (MSRA reductase) associated to H. vulgaris (Additional file 2: Table S19; Additional file 1: Figure S28). Finally, antibiotic resistance genes and virulence factors also show frequent phylogenetic and host-specific signals (Additional file 2: Tables S18 and S19; Additional file 1: Figure S28).

**Discussion**

Despite the great number of metagenomic studies published to date, which range in their focus on technical, analytical, or biological aspects, our study represents a unique contribution given its breadth of different host samples analyzed with a panel of standardized methods. In particular, the trade-offs between 16S rRNA gene amplicon and shotgun sequencing concerning amplification bias, functional information, and both monetary and computational costs warrant careful consideration when designing research projects. While 16S rRNA gene amplicon-based analyses are subject to considerable skepticism and criticism, we demonstrate that in many aspects similar, if not superior characterization of bacterial communities is achieved by these methods. We also show, however, that important insight can be gained through the combination of taxonomic and functional profiling, and that imputation-based functional profiles significantly differ from actual profiles. Our findings thus provide a guide for selecting an appropriate methodology for metagenomic analyses across a variety of metazoans. Finally, these data provide novel insight into the broad-scale evolution of host-associated bacterial communities, which can be viewed as particularly reliable given the repeatability of observations (e.g., differences between aquatic and terrestrial hosts, indicator taxa) across methods.

Given the concerns regarding the accuracy of 16S rRNA gene amplicon sequencing, other studies such as that of Gohl et al. [8] performed systematic comparisons of different library preparation methods and found superior results for a two-step amplification procedure. This method offers the additional advantage that one panel of adapter/barcode sequences can be combined with any number of different primers. Our first analyses were based on a standard mock community including Gram-positive and Gram-negative bacteria from the Bacilli and Gamma-Proteobacteria (eight species), as well as two fungi, which did not support an improvement of performance based on the two-step protocol. However, a number of changes were made to the Gohl et al. [8] protocol to adapt it to our lab procedures (e.g., larger reaction volumes, polymerase, variable region, heterogeneity spacers) that may contribute to these discrepancies, in addition to our different and diverse set of samples and other factors with potential influence on the performance of amplicon sequencing [6–8, 26–28]. The complexity of the mock community, i.e., the number of taxa, distribution, and phylogenetic breadth, may also have an influence on the discovery of clear trends in amplification biases or detection limits for certain taxonomic groups [29]. Thus, the even and phylogenetically shallow mock community in our study may be less suited than the staggered and diverse mixtures used in other studies [8] but still provides valuable information on repeatability, primer biases, and accuracy [29]. Nonetheless, when applied to our range of complex host-associated communities, we also found that significant differences in most parameters were due to the variable region rather than amplification method, and in many cases, biological signals were either improved or limited to the one-step protocol. Thus, in combination with the less complex laboratory procedures associated with the one-step protocol, we would generally recommend this procedure over two-step protocols.

Additional sources of variation influencing the outcome of our 16S rRNA gene amplicon-based community profiling are nucleic acid extraction procedures and the bioinformatic pipelines we employed. For the former, extraction procedures differed between host species due to
specific optimizations required for individual host species. Thus, certain differences in taxonomic and functional composition may be influenced by the specific protocols employed, as observed elsewhere [30]. Differences in the latter range from trimming and merging to clustering and classification, which are stringent and incorporate more reliable de novo clustering algorithms [31] as well as different classification databases [32]. Heterogeneity among the different amplicon approaches is however smaller than the differences between the amplicon and shotgun methods, as observed in other benchmarking studies [27]. Differences between shotgun approaches have been investigated in detail and also yield varying performances among classifiers, but in general, find a comparatively high performance of MEGAN-based approaches [9, 33, 34], which we also confirm in our study.

Given the limited number of studies that have compared imputed- and shotgun-derived functional repertoires [5, 35], our study also provides important additional insights. As imputation by definition is data-dependent, the differential performance and prediction among hosts in our study may in large part be explained by the amount of bacteria isolated, sequenced, and deposited (16S rRNA or genome) from these hosts or their respective environments. This seems to be most critical for the aquatic hosts. Furthermore, we observe a clear effect of variable region on the prediction performance, which is most obvious based on the mock community. The PICRUSt algorithm was developed and tested using primers targeting V3 V4 16S rRNA, and thus optimization of the imputation algorithm might be biased towards this target over the V1 V2 variable region. Although these performance differences, in particular the bias towards model organisms compared to less characterized communities (e.g., hypersaline microbial mats), were previously shown [5], our study provides additional, experimentally validated guidelines for a number of novel host taxa.

Interestingly, the strongest correspondence between bacterial community similarity and host genetic distance was detected at the bacterial order level for most of the employed methods. This may on the one hand reflect the deep phylogenetic relationships between our host taxa, such that turnover of bacterial taxa erodes phyllosymbiosis over time [19, 20]. On the other hand, some of the more striking observations made among our host taxa are the differences between aquatic and terrestrial hosts, both at the level of alpha and beta diversity. Based on a molecular clock for the 16S rRNA gene of roughly 1% divergence per 50 million years [36], bacterial order level divergence corresponds well with the timing of animal terrestrialization (425–500 MYA) [37, 38]. Although evolutionary rates can widely vary among bacteria species [39], other studies of individual gut microbial lineages such as the Enterococci indicate that animal terrestrialization was indeed a likely driver of diversification [40]. Specifically, the changing availability of carbohydrates in the host gut can be seen as a main driver of this diversification, which is consistent with the association of CAZY-based functional repertoires correlating to phylogenetic distance in our data set [19, 41].

In contrast to the patterns observed based on 16S rRNA gene amplicon-based profiles, the differentiation of bacterial communities according to host habitat was less pronounced based on functional genomic repertoires. This raises the possibility that the colonization of land by ancient animals required the acquisition of new, land-adapted bacterial lineages to perform some of the same ancestral functions. The overall observation of increased beta diversity among terrestrial compared to aquatic hosts (Additional file 1: Figure S19) could in part reflect differential acquisition among host lineages after colonizing land, although dispersal in the aquatic environment may on the other hand act as a greater homogenizing factor among aquatic hosts. The stronger correspondence between bacterial community and host phylogenetic distance among terrestrial hosts is also generally consistent with this hypothesis. However, the higher alpha diversity and the slightly lower correspondence with the phylogenetic patterns in aquatic hosts may also indicate a higher influence of environmental bacteria or a lack of physiological control over bacterial communities.

Bacterial taxa and functions involved in carbohydrate utilization were among the most notable associations to individual hosts, groups of hosts, and/or host phylogenetic relationships. Taxa such as Bacteroidales, Ruminococcaceae/Ruminococcales, and Clostridia associated to humans and/or mice include members known for a mucusal lifestyle, and these hosts also display the most diverse and abundant repertoire of carbohydrate-active enzymes (particularly glycosylhydrolases) in their microbiome. Other examples include sialidases, esterases, and fucosyltransferases, as well as different extracellular structures that appear to be specific to aquatic hosts, indicating differences in mucus and glycan composition according to this host environment. Glycan structures provide a direct link between the microbial community and the host via attachment, nutrition, and communication [42, 43], and the composition of mucin and glycan structures themselves show strong evolutionary patterns and are distinct among taxonomic groups [41]. Thus, a high diversity of glycan structures within and between hosts may determine the specific sets carbohydrate-facilitating enzymes of the respective microbial communities.

In addition to the bacterial carbohydrate hydrolases that digest surrounding host and dietary carbohydrates,
we also identified a number of glycosyltransferases associated with capsular polysaccharide synthesis (Additional file 2: Tables S19 and S20). This type of glycosylation is an important facilitator for host association and survival [44] and plays a crucial role in infections [45] in mutualists and pathogens alike [44, 46]. Thus, capsular and excreted glycan structures are important for the successful colonization and persistence in different environments [47, 48] and host organisms [44, 48].

Conclusions
In summary, the systematic comparison of five different metagenomic sequencing methods applied to ten different holobiont yielded a number of novel technical and biological insights. Although important exceptions will exist, we demonstrate that broad-scale biological patterns are largely consistent across these varying methods. As many aspects of differential performance in our study are host-specific (more detailed description of individual hosts can be found in Additional file 1), future development and benchmarking analyses would also benefit from including a range of different host/environmental samples.

Methods
DNA extraction and 16S rRNA gene amplicon sequencing
Protocols for each host type are described in Additional file 1: Figures S18–S28. Each library (16S rRNA gene amplicon, shotgun) included at least one mock Salmonella enterica (9.0%), Enterococcus faecalis (14.6%), Pseudomonas aeruginosa (10.4%), Staphylococcus aureus (14.6%), Staphylococcus epidermidis (13.2%), Bacillus subtilis (13.2%) and two fungi (Saccharomyces cerevisiae (1.6%), Cryptococcus neoformans (1.8%)).

The 16S rRNA gene was amplified using uniquely barcoded primers flanking the V1 and V2 hypervariable regions (27F–338R) and V3 V4 hypervariable regions (515F–806R) with fused MiSeq adapters and heterogeneity spacers in a 25-μl PCR [28]. For the traditional one-step PCR protocol, we used 4 μl of each forward and reverse primer (0.28 μM), 0.5 μl dNTPs (200 μM each), 0.25 μl Phusion Hot Start II High-Fidelity DNA Polymerase (0.5 U), 5 μl of HF buffer (Thermo Fisher Scientific, Inc., Waltham, MA, USA) and 1 μl of undiluted DNA. PCRs were conducted with the following cycling conditions in a 25-μl PCR reaction (98 °C, 30 s; 25 × [98 °C, 10 s; 55 °C, 30 s; 72 °C, 60 s]; 72 °C, 10 min; 10 °C, infinity) [8]. Following the PCR, the product was diluted 1:10 and 5 μl were used in an additional reaction of 10 μl (98 °C, 30 s; 10 × [98 °C, 9 s; 55 °C, 30 s; 72 °C, 60 s]; 72 °C, 10 min; 10 °C, infinity) utilizing the Nextera adapter overhangs to ligate the Illumina adapter sequence and individual MIDs to the amplicons, following the manufacturer’s instructions. The PCR protocol we used was 1 μl of each forward and reverse primer (5 μM), 0.3 μl dNTPs (10 μM), 0.2 μl Phusion Hot Start II High-Fidelity DNA Polymerase (2 μl), 2 μl of 5 × HF buffer (Thermo Fisher Scientific, Inc., Waltham, MA, USA), and 5 μl of the diluted PCR product. The concentration of the amplicons was estimated using a Gel Doc™ XR+ System coupled with Image Lab™ Software (BioRad, Hercules, CA USA) with 3 μl of O’GeneRulerTM 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, MA, USA) as the internal standard for band intensity measurement. The samples of individual gels were pooled into approximately equimolar sub-pools as indicated by band intensity and measured with the Qubit dsDNA br Assay Kit (Life Technologies GmbH, Darmstadt, Germany). Sub-pools were mixed in an equimolar fashion and stored at −20 °C until sequencing.

Library preparation for shotgun sequencing was performed using the NexteraXT kit (Illumina) for fragmentation and multiplexing of input DNA following the manufacturer’s instructions. Amplicon sequencing was performed on the Illumina MiSeq platform with v3 chemistry (2 × 300 cycle kit), while shotgun sequencing was performed on an Illumina NextSeq 500 platform via 2 × 150 bp Mid Output Kit at the IKMB Sequencing Center (CAU Kiel, Germany).

Amplicon analysis
The respective V1 V2 and V3 V4 PCR primer sequences were removed from the sequencing data using cutadapt (v.1.8.3) [49]. Sequence data in FastQ format was quality trimmed using sickle (v1.33) in paired-end mode with default settings and removing sequences dropping below 100 bp after trimming [50]. Forward and reverse read were merged into a single amplicon read using VSEARCH allowing fragments with a length of 280–350 bp for V1 V2 and 350–500 bp for V3 V4 amplicons [51]. Sequence data was quality controlled using fastq_quality_filter (FastX Toolkit) retaining sequences with no more than 5% of per-base quality values below 30 and subsequently with VSEARCH discarding sequences with more than one expected error [51, 52]. Reference-guided chimera removal was performed using the gold.fa
reference in VSEARCH (v2.4.3). The UTAX algorithm was used for a fast classification of the sequence data in order to remove sequences not assigned to the domains Bacteria or Archaea and exclude amplicon fragments from Chloroplasts [53]. Notably, only a total of 15 sequences were assigned to the domain Archaea, all found in two samples of human feces, accounting for less than 0.1% of the clean reads in these samples. The entire cleaned sequence data was concatenated into a single file and dereplicated and processed with VSEARCH for OTU picking using the UCLUST algorithm [54] using a 97% similarity threshold. OTUs were again checked for chimeric sequences, now using the de novo implementation of the UCHIME algorithm in VSEARCH [51, 54, 55]. All clean sequence data of the samples were mapped back to the cleaned OTU sequences using VSEARCH. OTU sequences and clean sequences mapping to the OTUs were taxonomically annotated using the RDP classifier algorithm with the RDP training set 14 [56, 57]. Sequence data were normalized by selecting 10,000 random sequences per sample. Taxon-by-sample abundance tables were created for all taxonomic levels from Phylum to Genus, as well as for OTUs.

PICRUSt functional imputations
Species-level OTUs (97% similarity threshold) were further classified using the GreenGenes (August 2013) database [58] via RDP classifier as implemented in mothur (v1.39.5) and merged with the abundances into a biome file which was uploaded to the Galaxy PICRUSt v1.1.1 pipeline (http://galaxy.morganlangille.com/) to derive functional imputations (COG predictions) [5]. To achieve accurate functional predictions, samples with NSTI ≤ 0.15 (weighted Nearest Sequenced Taxon Index) were pruned from the data set, as recommended by the developers.

Shotgun sequencing
Raw demultiplexed sequences were trimmed via Trimmomatic (v0.36) for low-quality regions with a minimum length of 50 bp as well as for adaptor and remaining MID sequences [59]. After trimming reads were mapped to host-specific genome databases and ΦX with additional retention databases containing all fully sequenced bacterial and metagenomic genomes (5 September 2015) via DeconSeq (v0.4.3) [60]. Single and paired sequences were repaired using the BBTools (v37.28) repair function [61]. Combined sequences were searched against the non-redundant NCBI database (28 July 2017) via DIAMOND [62] with (E value cutoff 0.001, v0.8.28) and MEGAN [13] classifying hits by functions (EggNOG—October 2016) and taxa (May 2017) (v6.6.1). For assemblies of single samples, we used metaSPADES [63] (v3.9.1) using paired reads in addition to unpaired reads left from the previous steps. PROKKA (v1.12) was used for gene calling and initial genome annotation [64] using the metagenome option with additional identifying rRNA and snRNA via barnap, ARAGORN [65], and Infernal [66]. ORFs were further annotated via EggNOG annotation via HMMER models implemented in the EggNOG-mapper (v0.12.7) [16, 67], CAZY database via dbCAN (v5, July 24, 2016), and HMMER3 [17, 68]. Gene abundances were derived from mapping the all reads back to the predicted ORF via bowtie2 (v2.2.6) [69] and calculated TPM (transcripts per kilobase million) via SamTools (v1.5) [70].

18S rRNA genes were obtained from NCBI GeneBank and aligned via ClustalW (v1.4) [71] for host tree construction, which includes A. aerophoba (gi:51095211, AY5917991), M. leidyi (gi:145177703, AF2937001), H. vulgaris (gi:761889987, JN5940542), A. aurita (gi:1470050, AY0392081), N. vectensis (gi:138977746, AF2543821), T. aestivum (gi:15982656, AY0490401), M. musculus (gi:374088232, NR_0032783), H. sapiens (gi:36162, X032051), D. melanogaster (gi:939630477, NR_1335591), and C. elegans (gi:30525807, AY2681171). Phylogenetic distance was calculated via DNADIST (v3.5c) [72] and a maximum likelihood tree was constructed via FastTree v2.1 CAT+ model [73]. Accuracy was improved via increased minimum evolution rounds for initial tree search (–spr 4), more exhaustive tree search (–mlacc 2), and a slow initial tree search (–slownni).

Statistical analysis
Statistical analyses were carried out via R (v3.4.3) [74]. Alpha diversity indices (richness, Shannon-Weaver index) and beta diversity metrics based on the shared presence (Jaccard distance) or abundance (Bray-Curtis distance) of taxa were calculated in the vegan package [75] and ordinated via Principal Coordinate Analysis (PCoA, avoiding negative eigenvalues), or via non-metric multidimensional scaling (NMDS) using a maximum of 10,000 random starts to obtain a minimally stressed configuration in three dimensions. Clusters were fit via an iterative process (10,000 permutations) and tested for separation by direct gradient analysis via distance-based redundancy analyses and permutative ANOVA (10,000 permutations) [76, 77]. Univariate analyses were carried out with approximate Wilcoxon/Kruskal tests as implemented in coin [78] (10,000 permutations). Procrustes tests were used to relate pairwise community distances based on either different data sources such as functional repertoires or taxonomic composition, as well as phylogenetic distances [21, 79]. Moran’s I eigenvector technique was employed to correlate bacterial community members and their functions to phylogenetic divergence, as implemented in ape (10,000 permutations) [22, 80].
Indicator species analysis, employing the generalized indicator value ($IndVal_g$), was used to assess the predictive value of a taxon for each respective host phenotype/category as implemented in indispecies [15]. Linear mixed models, as implemented in nlme were used to compare the influence of amplification method or variable region without the influence of the organism of origin [81]. We employed the Hommel and Benjamini-Yekutieli adjustment of $P$ values when advised [82, 83].

**Additional files**

- **Additional file 1**: Supplementary Materials. (PDF 6900 kb)
- **Additional file 2**: Supplementary Tables. (ZIP 1765 kb)

**Acknowledgements**

We thank Katja Cloppenborg-Schmidt, Melanie Vollstedt, and Dr. Sven Künzel for the excellent assistance and help during the development of the project and their constant drive to improve its quality.

**Authors’ contributions**

PPa, PRo, AF, TB, and JFB conceived and designed research. PPa and MR performed data analyses. PPa, MR, BH, SD, and JFB interpreted the results and wrote the manuscript. PPa, MR, TD, KD, HD, SD, SF, JF, UHH, FAH, BH, MH, MU, CJ, KABK, DL, AR, TBRH, TR, RAS, HS, RS, FS, ES, NWB, PRo, AF, TB, and JFB generated and interpreted host-specific data and gave intellectual input. All authors read and approved the final manuscript.

**Funding**

This work was funded by the DFG Collaborative Research Centre (CRC) 1182 “Origin and Function of Metaorganisms” subproject Z3 and the Max-Planck-Society.

**Availability of data and materials**

Sequence and meta-data are accessible under the study identifier PRJEB30924 (“https://www.ebi.ac.uk/ena/”). Remaining DNA from non-human samples can be made available upon request. All human samples and information on their corresponding phenotypes have to be obtained from the PopGen Biobank Kiel (Schleswig-Holstein, Germany) through a Material Data Access Form. Information about the Material Data Access Form and how to apply can be found at “https://www.uksh.de/p2n/Information+for+Researchers.html”.

**Ethics approval and consent to participate**

Human samples

Study participants were randomly recruited from inhabitants of Schleswig-Holstein (Germany) who were recruited for the PopGen cohort. Five individuals from the PopGen biobank (Schleswig-Holstein, Germany) were randomly selected among the healthy and unmedicated individuals and included in the study without corresponding meta-information. Study participants collected fecal samples at home without conservation buffers in standard fecal tubes (sterile feces container 76 × 20 mm, Sarstedt) and shipped them immediately at room temperature or brought them to the collection center (within 24 h). Samples were stored at −80 °C until processing. Human feces (N = 4) were sampled and extracted following the procedures as described in Wang et al. 2016 [84]. A biopsy sample of the sigmoid colon was taken from a healthy control individual without macro- or microscopical inflammation (N = 1) and DNA was extracted as described in Rausch et al. [85]. Investigators were blinded to sample identities and written informed consent was obtained from all study participants before the study. All protocols were approved by the Ethics Committee of the Medical Faculty of Kiel and by the data protection officer of the University Hospital Schleswig-Holstein in accordance with the Declaration of Helsinki Principles.

Animal and plant samples

Wild-derived, hybrid mice were sacrificed according to the German animal welfare law and Federation of European Laboratory Animal Science Associations guidelines. Hybrid breeding stocks of wild-derived M. m. musculus × M. m. domesticus hybrids captured in 2008 are kept at the Max Planck Institute Plön (11th lab generation). The approval for mouse husbandry and experiment was obtained from the local veterinary office “Veterinärmärkte Kreis Plön” (Permit: 1401-144/PL0-004697). All samplings, including invertebrate and plant samples, were performed in concordance with the German animal welfare law and Federation of European Laboratory Animal Science Associations guidelines. Further details for each host type are provided in Additional file 1.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1. Evolutionary Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany. 2. Institute for Experimental Medicine, Kiel University, Kiel, Germany. 3. Department of Biology, Laboratory of Genomics and Molecular Biomedicine, University of Copenhagen, Copenhagen Ø, Denmark. 4. Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. 5. Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany. 6. Institute of General Microbiology, Kiel University, Kiel, Germany. 7. Department of Evolutionary Ecology and Genetics, Zoological Institute, Kiel University, Kiel, Germany. 8. Zoological Institute, Kiel University, Kiel, Germany. 9. Molecular Physiology, Zoological Institute, Kiel University, Kiel, Germany. 10. Marine Ecology, Research Unit Marine Symbioses, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany. 11. Kiel University, Kiel, Germany. 12. Marine Ecology, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany. 13. Environmental Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany. 14. Environmental Genomics, Botanical Institute, Kiel University, Kiel, Germany.

Received: 8 February 2019 Accepted: 23 August 2019

Published online: 14 September 2019

**References**

1. Bosch TCG, McFall-Ngai MJ. Metaorganisms as the new frontier. Zoology. 2011;114(4):185–90.
2. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Loso T, Douglas AE, Dubiller N, Eberl G, Fukami T, Gilbert SF, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci. 2013;110(9):3229–36.
3. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. Microb Ecol Health Dis. 2015;26(1):26191.
4. Morgan XC, Hutterower C. Chapter 12: human microbiome analysis. PLoS Comput Biol. 2012;8(12):e1002808.
5. Langille MGJ, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31(9):814–21.
6. Hiergeist A, Glaser J, Reichl U, Gessner A. Analyses of intestinal microbiota culture versus sequencing. ILAR J. 2015;56(2):228–40.
7. Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG. Primer and platform effects on 16S rRNA tag sequencing. Front Microbiol. 2015;6:771.
8. Gohl DM, Vangay P, Garbe J, McLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. Nat Biotechnol. 2016;34(9):942–9.
9. Walsh AM, Crispie F, O’Sullivan O, Finnegan L, Claesson MJ, Cotter PD. Species classifier choice is a key consideration when analysing low-complexity food microbiome data. Microbiome. 2016(4):50.
10. Faith JJ, Guruge JL, Chatbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, et al. The long-term stability of the human gut microbiota. Science. 2013;341(6141):1237439.
11. Jovel J, Patterson J, Wang W, Hotte N, O’Keefe S, Mitchell T, Perry T, Kao D, Mason AL, Madsen KL, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol. 2016;7(1459):459.
62. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59–60.
63. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. metaSPAdes: a new versatile de novo metagenomics assembler. arXiv preprint arXiv: 1604.03071. 2016.
64. Seemann T. Prokka rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
65. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–6.
66. Kolbe DL, Eddy SR. Fast filtering for RNA homology search. Bioinformatics. 2011;27(22):3102–9.
67. Huerta-Cepas J, Forslund K, Coelho LP, Saklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34(8):2115–22.
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(1):445–51.
69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
71. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through the inclusion of position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(13):4673–80.
72. Felsenstein J. DNADIST – Program to compute distance matrix from nucleotide sequences. 3.5c edn; 1993.
73. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.
74. Team RC. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing. 3.3.2 edn; 2016.
75. Oksanen J, Blanchet FG, Kindt R, Legendre P, O’Hara RB, Simpson GL, Sølymos P, Stevens MHH, Wagner H. vegan: Community Ecology Package 1. 17-6 edn: 2011 http://CRAN.R-project.org.
76. Legendre P, Anderson MJ. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monogr. 1999;69(1):1–24.
77. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology. 2001;26(1):32–46.
78. Hommer G. A stage wise rejective multiple test procedure based on a modified Bonferroni test. Biometrika. 1988;75(2):383–6.
79. Benjamin Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29(4):1165–88.

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