Computational and Structural Biotechnology Journal 19 (2021) 4658–4668

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Contribution of sample processing to gut microbiome analysis in the model Lepidoptera, silkworm *Bombyx mori*

Nan Zhang [a,1], Jintao He [a,1], Xiaoqiang Shen [a], Chao Sun [b], Abrar Muhammad [a], Yongqi Shao [a,c,*]

[a] Max Planck Partner Group, Institute of Sericulture and Apiculture, College of Animal Sciences, Zhejiang University, Hangzhou, China
[b] Analysis Center of Agrobiology and Environmental Sciences, Zhejiang University, Hangzhou, China
[c] Key Laboratory for Molecular Animal Nutrition, Ministry of Education, China

ARTICLE INFO

Article history:
Received 25 May 2021
Received in revised form 12 August 2021
Accepted 12 August 2021
Available online 17 August 2021

Keywords:
Gut microbiota
Sample processing
metagenomic DNA extraction
Microbial informatics
Sources of variation
Technical variability
Lepidoptera

ABSTRACT

Microbes that live inside insects play various roles in host biology, ranging from nutrient supplementation to host defense. Although Lepidoptera (butterflies and moths) are one of the most diverse insect taxa and important in natural ecosystems, their microbiotas are little-studied, and to understand their structure and function, it is necessary to identify potential factors that affect microbiome analysis. Using a model organism, the silkworm *Bombyx mori*, we investigated the effects of different sample types (whole gut, gut content, gut tissue, starvation, or frass) and metagenomic DNA extraction methodologies (small-scale versus large-scale) on the composition and diversity of the caterpillar gut microbial communities. High-throughput 16S rRNA gene sequencing and computational analysis of the resulting data unraveled that DNA extraction has a large effect on the outcome of metagenomic analysis: significant biases were observed in estimates of community diversity and in the ratio between Gram-positive and Gram-negative bacteria. Furthermore, bacterial communities differed significantly among sample types. The gut content and whole gut samples differed least, both had a higher percentage of *Enterococcus* and *Acinetobacter* species; whereas the frass and starvation samples differed substantially from the whole gut and were poor representatives of the gut microbiome. Thus, we recommend a small-scale DNA extraction methodology for sampling the whole gut under normal insect rearing conditions whenever possible, as this approach provides the most accurate assessment of the gut microbiome. Our study highlights that evaluation of the optimal sample-processing approach should be the first step taken to confidently assess the contributions of microbiota to Lepidoptera.

1. Introduction

Lepidoptera are the second largest order of insects and are widespread and widely recognizable in nature. Although butterflies and moths play an important role in ecosystems as pollinators and as prey in the food chain, their caterpillars are problematic in agriculture because their main food source is live plants. Due to the low nutrient content, indigestibility and toxicity of many plant tissues [1], herbivorous insects have developed numerous traits to overcome these dietary obstacles. While previous research has mostly focused on the counteradaptations rooted in the insect genome [2], microbial symbionts, particularly those inhabiting the gut, are becoming increasingly recognized as a significant player in insect-plant interactions [3–7].

In fact, various studies have shown the relative composition of lepidopteran caterpillar gut community. For example, both culture-dependent and culture-independent techniques revealed that *Enterococcus* species are especially common and consistent gut inhabitants found in a wide range of lepidopterans, including the wax moth *Galleria mellonella* [8–10], gypsy moth *Lymantria dispar* (Lepidoptera: Lymantriidae) [11], cotton leafworm *Spodoptera littoralis* (Lepidoptera: Noctuidae) [12], spring webworm *Ocnogyna loewii* (Lepidoptera: Erebidae) [13], spurge hawk-moth *Hyles euphorbiae* (Lepidoptera: Sphingidae) and Amaryllis borer *Brithys crini* (Lepidoptera: Noctuidae) [14]. In particular, enterococci are metabolically active inside the gut of the tobacco hornworm *Manduca sexta* (Lepidoptera: Sphingidae) and originate from eggs but not from host plants [15]. Metagenomic sequencing of the diamondback moth gut microbiome revealed three, highly-abundant

* Corresponding author at: Max Planck Partner Group, Institute of Sericulture and Apiculture, College of Animal Sciences, Zhejiang University, Hangzhou, China.
E-mail address: yshao@zju.edu.cn (Y. Shao).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.csbj.2021.08.020

bacteria (*Enterobacter cloacae*, *Enterobacter asburiae*, and *Carnobacterium maltaromaticum*) as key symbionts across the host life cycle [16]. Montagna et al. provide evidence of a core microbiota in the Indian meal moth *Plodia interpunctella* (Lepidoptera: Pyralidae) [17–18]. In addition to extracellular bacteria, endosymbionts are also widespread in Lepidoptera [13,19]. For example, *Wolbachia* could be detected in all leaf-mining moths of *Phyllonorycter blancardella* (Lepidoptera: Gracillariidae) collected from different locations and is believed to be the most highly abundant bacteria in this insect [20]. Although microbial communities associated with the gut of lepidopterans are attracting an increasing interest [21–22], systematic study of potential factors influencing microbiome analysis, as revealed by recent work in the human gut microbiome [23–24], is still necessary for Lepidoptera [25].

A multitude of gut microbiome studies have already uncovered significant heterogeneity between intestinal luminal, mucosal and fecal microbiota [24,26–27]. For instance, mucosa-associated bacteria significantly differ from the bacterial community recovered from feces, limiting the applicability of feces in fully assessing both the murine and human gut microbiome [24]. The gut of lepidopteran insects is quite simple without any specialized structures, and it accounts for the largest proportion of the body cavity [28]. Lepidopteran caterpillars often consume large amounts of plant material during development, thus their guts contain a large volume of plant biomass. To study the lepidopteran gut microbiome, several sampling procedures have been described in the literature for obtaining samples for the purpose of sequencing metagenomic DNA. For instance, the whole gut is homogenized and extracted directly [12–13,29–30]. Starvation prior to sampling commonly promotes the elimination of plant material from the gut [14]. The use of frass has also been reported in a variety of species [21]. Therefore, different sample types potentially affect microbiome analysis in the field.

More surprisingly, recent work has shown that DNA extraction methodology has the largest effect on the outcome of metagenomic analysis in human gut microbiome studies [23]. In particular, lepidopteran caterpillars grow fast, and larval mass and body volume increase dramatically. For instance, the weight of *M. sexta* increases more than 1,000-fold from the first to the fifth instar [31]. This significant change in body size could account for the sampling problem encountered across host development since early-instar caterpillars offer very limited sample material (at the milligram level per individual); thus, a small-scale DNA extraction methodology is most suitable for early instars. Late-instar caterpillars, however, provide a relatively large amount of sample material (at the gram level per individual), and a large-scale DNA extraction methodology is therefore more suitable. For small sample sizes, DNA purification methods with commercially available kits are widely used for next-generation sequencing-based analyses of microbial community composition, such as the MasterPure Complete DNA and RNA purification kit from Epicentre Biotechnologies (Illumina, USA) [32–35]. For large sample sizes, DNA can often be easily extracted by routine procedures, such as proteinase K digestion followed by phenol–chloroform-isoamyl alcohol (PCI) extraction and ethanol precipitation [36]. Use of inconsistent methods applied over the course of host development potentially further affects data sets.

As already highlighted above, sample processing must consider confounders when comparing multiple studies and data sets. Using the silkworm *Bombyx mori* (Lepidoptera: Bombycidae) as a research model in the present study, we conducted a systematic comparison of sample type and metagenomic DNA extraction methodology to identify whether these factors are also sources of variation in the gut microbial community of lepidopteran caterpillars. *B. mori*, having been domesticated over 10,000 years, is a powerful experimental model for both basic and applied research [37]. The inbred, standard strain p50 silkworms have a highly identical genetic background, and the interindividual variation in their gut microbiotas is also not significant [3], a phenomenon widely observed in Lepidoptera, especially in those individuals from the same population [12]. We used both culture-dependent techniques and culture-independent high-throughput sequencing of the 16S rRNA gene to compare bacterial load and community compositions. Evaluations of the influence of sampling and sample processing have important implications for the interpretation of lepidopteran gut microbiome studies.

2. Materials and methods

2.1. Insect rearing

Eggs of *B. mori* inbred strain p50 (*Dazao*) were provided by the Silkworm Germplasm Bank at the College of Animal Sciences, Zhejiang University, China, and hatched at 28 ℃ using standard protocols [38]. Newly hatched larvae were reared in groups (50 individuals per chamber) at 25 ± 1 ℃ and 70 ± 5% humidity on a 14:10 (light: dark) cycle, and were allowed to feed on fresh mulberry leaves *ad libitum*. Mulberry leaves were collected from a mulberry-planting field (30°18′N, 120°04′E) without further disinfection. Surface-disinfected gloves and lab coats were worn during the whole experimental process.

2.2. Sample collection and preparation

Forty silkworms were collected on the fourth day of 5th-instar larvae and stored immediately at −80 °C. Another 20 silkworms at the same stage were starved for 24 h in sterile Petri dishes, after which they were collected as starvation samples. And the fresh frass was frozen too at −80 °C after weighed. Before sample processing, all frozen larvae were first rinsed three times in sterile water, surface-sterilized in 70% ethanol for 30 s and washed three times again in sterile water to minimize external contamination. The whole gut (from proventriculus to rectum), gut content or gut tissue was dissected from each individual under aseptic conditions in Petri dishes, using sterile scissors and forceps according to Shao et al. [39]. The whole gut (including the gut tissue and the possible remaining food content) was also withdrawn from starved caterpillars. After being weighed, each dissected sample was transferred into a 2.0-mL lysing tube and then was homogenized directly via bead-beating without any suspension buffer using a Bertin lyser instrument (5000 rpm, 30 s). Soft and loose frass samples were homogenized by manual grinding with a pestle.

2.3. DNA extraction and assessment of extraction success

For DNA extraction, two commonly used methodologies were employed, one for small sample size and another for large-scale DNA extraction. Most commercial kits fit to small volume and we selected the MasterPure™ Complete DNA and RNA Purification Kit (Epicentre, Illumina), which is widely used based on both chemical and enzymatic lysis and seamlessly integrated into next-generation sequencing [3,29,40–41]. We followed all steps recommended by this kit, including the proteinase K and ribonuclease A (RNase A) treatment. An additional lysozyme incubation step (30 min at 37 ℃, 4 μL of 100 mg/mL lysozyme, Sigma-Aldrich, USA) was included prior to proteinase K digestion to break up Gram-positive bacterial cells [32]. The other one for large-scale DNA extraction was the classic phenol–chloroform-isoamyl alcohol (PCI) based extraction and ethanol precipitation. All sample types were extracted by both methods except frass (only the first method) due to small size (6.03 mg on average). Specifically, insect

tissue-free water "blanks" were processed with both DNA extraction methodologies and PCR amplification kits as the negative controls to detect reagent and environment contamination.

Twenty mg homogenate was first removed from each individual sample and extracted with the kit, using the manufacturer's recommended procedure; the rest part (182 mg on average, Sample mass, Table S1) was subject to a PCI extraction protocol according to Renshaw et al. [36]. Briefly, the homogenized sample was suspended in 25 mL of a solution containing 2 g of polyethylene glycol 2000, 62.5 mg of egg white lysozyme, and 5 mM Tris-hydrochloride (pH 8.0), and incubated in a 37 ℃ water bath for 2 h. After centrifugation (5 min, 8,000 rpm), the cells were resuspended in 12.5 mL of TES buffer (50 mM Tris-hydrochloride [pH 7.6], 20 mM sodium ethylenediaminetetraacetic acid, and 25% sucrose), treated with 12.5 μL of 20 mg/mL of RNase A for 60 min at 37 ℃, and then lysed with heat at 60 ℃ for 2 h in the presence of 20 μL of 50 mg/mL of proteinase K and 1.7% sodium dodecyl sulphate. After that, 2.13 mL of 5 M sodium chloride solution and 1.7 mL of cetyltrimethyl ammonium bromide (CTAB) buffer (0.7 M NaCl and 10% (w/v) CTAB) was added and incubated in a 65 ℃ water bath for 10 min. Then the samples were centrifuged at 8000 rpm for 20 min and the cleared supernatants were transferred to fresh tubes to be extracted. The same volume of PCI mixture in a ratio of 25:24:1 was added to the lysate and samples were vortexed for 5 s. Tubes were centrifuged at 8000 rpm for 20 min and the aqueous layer was transferred to a fresh 50-mL Falcon polypropylene centrifuge tube. The extraction procedure was repeated once, and the aqueous layer was transferred again to a fresh tube. Two volumes of 100% ice-cold ethanol were added to the aqueous supernatant and the samples were precipitated at −20 ℃ overnight. The precipitate was pelleted by centrifugation at 8000 rpm for 20 min, and the liquid was decanted. Then the pellet was washed twice with 10 mL of 70% ethanol at room temperature. The ethanol was removed by centrifugation at 8000 rpm for 15 min. Pellets were dried on a laminar flow clean bench at room temperature until no visible liquid remained; and finally pellets were redissolved in sterile Millipore water. DNA concentration and purity was measured by Biodropsis BD-2000 instrument (OSTC, China), and the integrity was determined by electrophoresis on a 1% (w/v) agarose gel. The successful extraction of bacterial DNA was verified using PCR assays with universal eubacterial 16S rRNA primers (27f and 1492r). Subsequently, the extracted DNA was used for Illumina sequencing.

2.4. Bacterial enumeration

The gut was dissected from the insect as previously described [39]. Each sample was homogenized in the sterile PBS buffer and diluted serially ranging from 1×10^{-1} to 1×10^{-6}. Considering the different nutritional needs of gut bacteria, two types of media (LB and Brain and Heart Infusion (BHI)) were used for culturing. A 50-μL aliquot was spread onto agar plate. The colony forming units (CFUs) were counted on each plate after a 48 h incubation at 30 ℃.

2.5. Amplification and Illumina sequencing of targeted 16S rRNA gene

DNA extracts from different sample types were amplified with a primer pair specific for the V3-V4 region of the 16S rRNA gene (338f: ACTCCTACGGGAGGCAGCAG, 806r: GGACTACHVGGGTWTCTAAT) respectively. The PCR reaction was performed in triplicate for each sample and the sequencing primers contained the adapter and barcode sequences. We included appropriate negative controls (blank samples) at all steps in PCR reactions. For each reaction, 20 μL of the mix was prepared, containing 4 μL FastPfu reaction buffer (TransGen, China), 2 μL 2.5 mM dNTPs, 0.8 μL 5 μM of each primer, 0.4 μL FastPfu Polymerase, 0.2 μL BSA, and 10 ng of template DNA. The PCR reaction involved a single denaturation step at 95 °C for 3 min, followed by 27 cycles of 95 °C for 30 s, 53 °C for 30 s, 72 °C for 45 s, and finished after a final extension at 72 °C for 10 min. The triplicate reaction products were pooled and run on a 2% (w/v) agarose gel. Gel fragment of correct size was excised and purified with an AxyPrep DNA gel extraction kit (Axygen, USA). After quantified by Quantifluor dsDNA system (Promega, USA), products were calculated into equal amount and mixed for Illumina MiSeq (Illumina, USA) paired-end sequencing performed by a certified sequencing provider (Majorbio, China). 66 DNA samples were sequenced successfully: 18, 24 and 24 samples belonging to the experiment "Methodology reproducibility and variation" (for each method (the kit or PCI), three technical replicates (each 20 mg) were withdrawn from the well-mixed whole gut material of each single 5th-instar silkworm individual (biological replicate)), "Microbiota similarities and differences between sample types under small-scale extraction conditions" (a single individual (biological replicate) for each sample) and "Microbiota similarities and differences between sample types under large-scale extraction conditions" (a single individual (biological replicate) for each sample), respectively.

2.6. Data processing and statistical analyses

Demultiplexed sequences were merged with FLASH (v1.2.11) [42] and quality filtered with fastp (v0.19.6) [43] according to the criteria previously described [44]. Usearch (v8.1.1861) [45] algorithm was used for removal of the chimeric sequences and OTU clustering with a similarity threshold of 97%. After removing singletons, the most abundant sequence in each OTU was defined as the representative sequence, which was assigned against the SILVA database (v132) [46], with minimum confidence 0.8. As an additional quality control measure, OTUs classified as chloroplasts or mitochondria and those that could not be assigned to a specific phylum were removed prior to further analyses [47]. The presence of potential contaminant sequences was identified (Table S2) using decontam package (v. 1.2.1) based on *isContaminant* function at the default classification threshold [48]. The non-contaminant sequence data were randomly rarefied to the minimum number (2416) sequences in all sequenced samples to account for the variable sequencing depth obtained and a normalized OTU table was generated. Mothur (v1.37.6) analysis of data sets was conducted with the transferred OTU abundance table by the make.shared command [49]. The DivNet R package was used to estimate diversity according to Willis, A.D., and Martin, B.D. [50].

Following initial processing, Principal Coordinate Analysis (PCoA) was performed based on the weighted UniFrac dissimilarities, and P-values were calculated by the permutational multivariate analysis of variances (PERMANOVA) in PAST (v3.21) using 9999 permutations [51]. Heatmaps and Venn plots were generated based on the relative abundance of OTUs at the genus level, visualized by using the R package "ggplot2" and "VennDiagram". Correlation heatmaps were based on the Spearman distance calculated by the relative abundance of each OTU at the genus level. Variation of total DNA quantity, 16S rRNA gene copies and alpha diversity indices were calculated by Student's *t*-test or Wilcoxon rank-sum test using SPSS software (v20.0, IBM Corporation, USA). Normal distributions of the data were checked with the Shapiro–Wilk test and homoscedasticity of variances was analyzed using either Bartlett's or the Fligner–Killeen test.

2.7. Data accessibility

The sequencing data are available through the NCBI SRA archive, accession number: SRP100894 and SRX3759742.

3. Results

3.1. Study design and quality control of extracted DNA

As in many other lepidopteran herbivores, there was significant variation in body size across silkworm development. The final (5th) instar larva of *B. mori* reared on native plant diet was much larger than other instar larva (Fig. 1a). Fig. 1a also shows a large amount of plant material inside the tube-like digestive tract, which is the largest part of the body of a silkworm. We compared the bacterial communities present across the five most commonly used sample types (Fig. 1b) and between DNA extraction methods (small-scale vs. large-scale). DNA extracted directly from the whole gut (W) of 5th-instar larvae represents total gut microbial biomass.

Almost all molecular techniques to describe and quantify the microbiota require extraction of DNA from samples as a first step. Thus, quality control of extracted DNA, in terms of DNA integrity and quantity, was first considered. High-molecular-weight DNA was recovered from both extraction methodologies, and the DNA quality was shown to be reproducible across individuals (Fig. S1). The size of most DNA fragments obtained was above 15 kb. The extracted DNA was sufficiently pure to be used directly as a template for PCR and was therefore suitable for high-throughput sequencing. However, despite the use of more gut materials, the PCI protocol did not appear to yield more DNA (Fig. 2a). Overall, DNA extracted using the kit gave an average yield of 40.2 μg (range 8.3–73.5 μg) per sample, whereas only 14.4 μg (range 6.6–28.9 μg) per sample was obtained using the PCI protocol. In particular, there was a significant difference between the two methodologies in the gut tissue and starvation sample yields, with P values of 0.003 and 0.001, respectively (Student's *t* test). Furthermore, the kit procedure provided substantial improvements in extraction efficiency (Fig. 2b). An obviously increased extraction efficiency of DNA (μg DNA extracted per mg sample) was observed in the whole gut ($P < 0.001$), gut tissue ($P = 0.001$) and starvation ($P < 0.001$) samples.

3.2. Methodology reproducibility and variation

Since DNA quantity had already shown a between-method difference, we next investigated the influence of DNA extraction methodology on the outcome of metagenomic analyses. We compared gut microbiota structure by directly sequencing DNA extracted from the same 5th-instar individual caterpillar with the two different methodologies to reduce biological sources of variation (Fig. 3). For each method, three test portions (20 mg per sample) were withdrawn from the well-mixed whole gut material of each individual, and DNA was extracted by either the kit or the PCI procedure as described in materials and methods (Fig. 3a). Under standard laboratory conditions, the commercial kit-based procedure extracted DNA that is sufficiently pure to be used as a template for PCR within an hour. By contrast, the PCI approach involved more hands-on time because of the additional matrix removal step.

The bacterial community structure was determined by Illumina MiSeq sequencing of the 16S rRNA gene. Despite the two DNA extraction procedures shared the same 23 genera (Fig. 3b), more OTUs were found in the case of the kit-based procedure, probably due to a significantly higher proportion of Gram-positive bacteria being detected (Fig. 3c).

In terms of technical sources of variation, we assessed measurement reproducibility through technical replication. The taxonomic composition of microbiota among technical replications of the same individual was consistent in the results from the kit method, with samples collected from the same individual clustering together; however, the taxonomic composition of technical replicates derived from the PCI method varied widely, with samples collected from the same individual clustering separately (Fig. 3d). This pattern was apparent in all biological samples tested, suggesting that technical variation within the PCI methodology was a considerable factor in lepidopteran microbiome analysis. Notably, samples extracted with the same methodology, regardless of the individual silkworms they originated from, all clustered into the same clade (kit or PCI-based); thus, between-methodology variation was greater than the variation produced by biological effects. The heatmap of the most proportionally abundant bacterial taxa showed that the hard-to-extract Gram-positive bacteria, such as *Enterococcus*, were heavily underestimated when compared to the mean abundances across the nine replicates, from 55.7% in the kit method to 7.2% in the PCI method, while Gram-negative bacteria, such as *Ralstonia*, were largely overrepresented in the PCI method, indicating that this procedure probably was only effective in recovering DNA from a variety of Gram-negative bacteria (Fig. 3d).

Altogether, major intermethod differences in the composition of whole gut microbiota were evident as determined by molecular biological techniques. DNA extraction also has a large effect on the outcome of metagenomic analysis of the lepidopteran microbiome.

3.3. Bacterial load and overall community diversity across sample types

The culture-based approach revealed that abundant bacteria colonized the silkworm (Fig. 4). In the whole gut, the cultivatable bacterial count on LB agar was found to be 5.48×10^8 CFU/g sample on average. The number of bacteria decreased in the gut content (8.06×10^7 CFU/g sample), gut tissue (3.36×10^7 CFU/g sample), and starved caterpillars (1.12×10^8 CFU/g sample). The cultivatable colonies were approximately 8.48×10^8 CFU/g in the frass. A similar trend on BHI agar was observed across all five sample types (whole gut, 5.52×10^8 CFU/g; gut content, 1.03×10^8 CFU/g; gut tissue, 6.73×10^7 CFU/g; starved caterpillar, 1.04×10^8 CFU/g; frass, 9.72×10^8 CFU/g).

DNA extracted from the herbivore gut potentially includes microbial DNA, host (*B. mori*) DNA and plant diet (*Morus alba*) DNA. When sampling bacterial communities in lepidopteran herbivores, chloroplast sequence contamination in 16S analyses can be particularly problematic since lepidopteran caterpillars consume large amounts of plant material [34]. As expected, compared to gut tissue and starvation sample types, the whole gut, gut content and frass samples, which contained large amounts of plant materials, also had a relatively high representation of sequences related to chloroplasts (Fig. 5a).

By using the extraction kit, we found that the whole gut was associated with the highest number of phylotypes (49 OTUs on average), reflecting the true extent of community diversity (Fig. 5b). The total number of OTUs decreased in the frass samples (only 29 OTUs on average). Notably, lower OTU numbers were recovered with the PCI procedure than with the kit-based procedure in all sample types, especially in the whole gut and gut content samples, with P values of <0.0001 and 0.0085, respectively (Student's *t* test). In addition, intersample differences were not evident as determined by PCI extraction. Consistently, the estimated Shannon index also showed that bacterial biodiversity present in the whole gut (2.75) was higher than that present in other sample types (Fig. 5c). The kit method reliably detected more species from the whole gut ($P = 0.0013$). Notably, starvation and frass samples exhibited significantly lower diversity, with mean values of the Shannon index at 1.12 and 0.90 ($P = 0.0221$ and 0.0306,

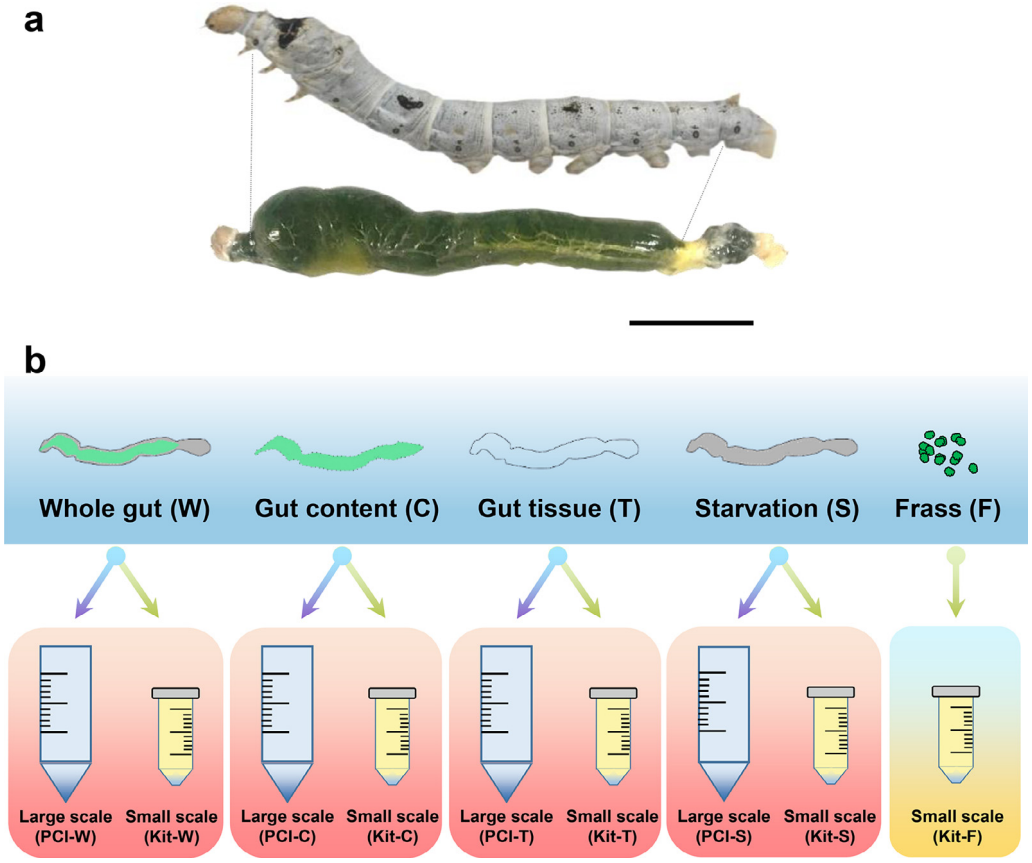

Fig. 1. An overview of the experimental design. (a) The final (5th) instar silkworm and its gut. Scale bar, 1 cm. (b) Outline of the sampling setup. Composition and variability of the silkworm gut microbiotas are compared across all sample types (W, whole gut; C, gut content; T, gut tissue; S, starvation; F, frass) by using both large-scale (PCI) and small-scale (Kit) metagenomic DNA extraction methods. A small (20 mg per sample) and a large (182 mg on average) test portion were collected for each individual and extracted with the two methods respectively; the frass was only extracted using the Kit method due to its small size.

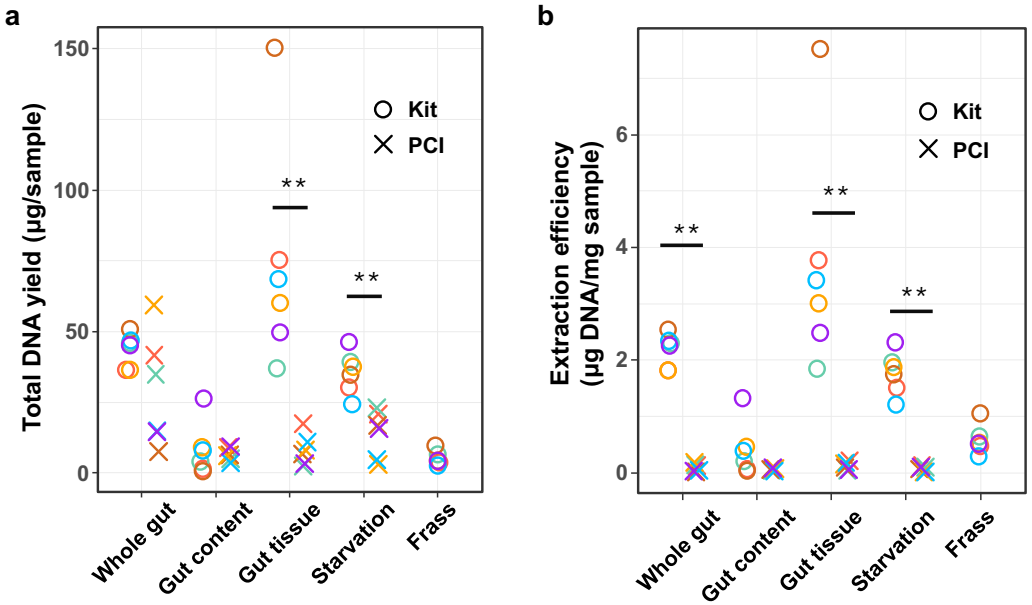

Fig. 2. Quantity of DNA extracted from different sample types by using both small-scale (Kit) and large-scale (PCI) methods. (a) Total DNA yield per sample (for Kit, 20 mg sample used; for PCI, 182 mg sample used on average). (b) Characterization of extraction efficiency by method type. Each symbol represents a single individual (biological replicate). **, $P < 0.01$, Wilcoxon rank-sum test.

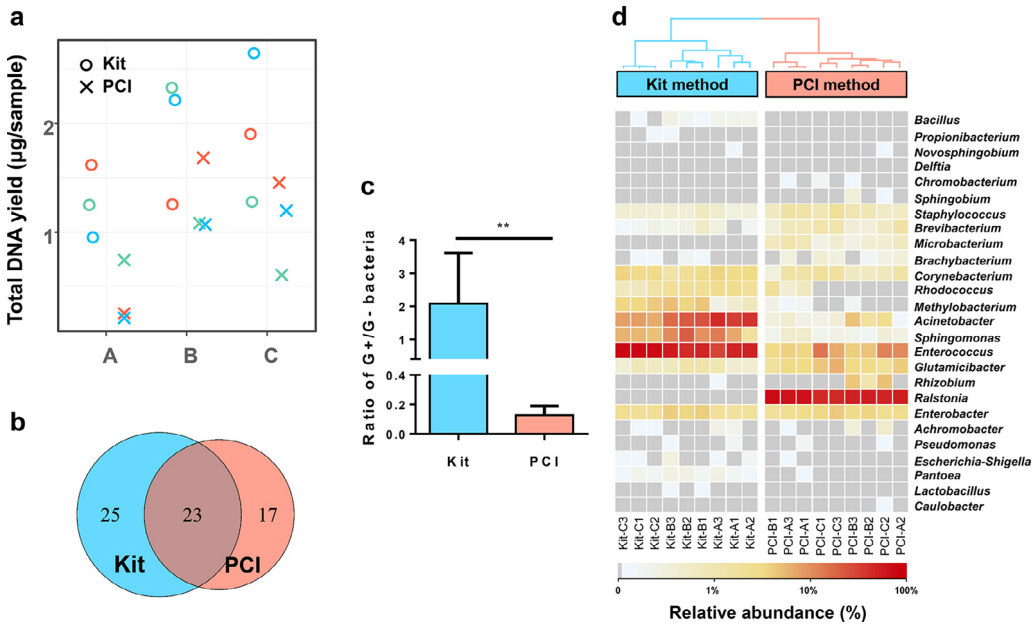

Fig. 3. Effect of metagenomic DNA extraction methodology on microbiota structure. (a) The yield of DNA extracted using the kit and PCI method across biological replicates (independent 5th-instar silkworm individuals). A, B and C represent different individuals (biological replicate); each symbol represents technical replication (20 mg subsample originated from the homogenized whole gut of a single individual). (b) Venn diagram showing the numbers of genera shared or unique between the kit and PCI-based data sets. (c) Recovery of the hard-to-extract Gram-positive bacteria (G+) between the two methodologies. **, significance (P = 0.0015, Student's *t* test). (d) Heatmap of major taxa identified. Cluster analysis used the Bray-Curtis distance and complete-linkage algorithm. Genus-specific abundance variation shows that biases are consistent across the two extraction methods. A, B and C represent different individuals (biological replicate); 1, 2 and 3 represent technical replication of a single individual.

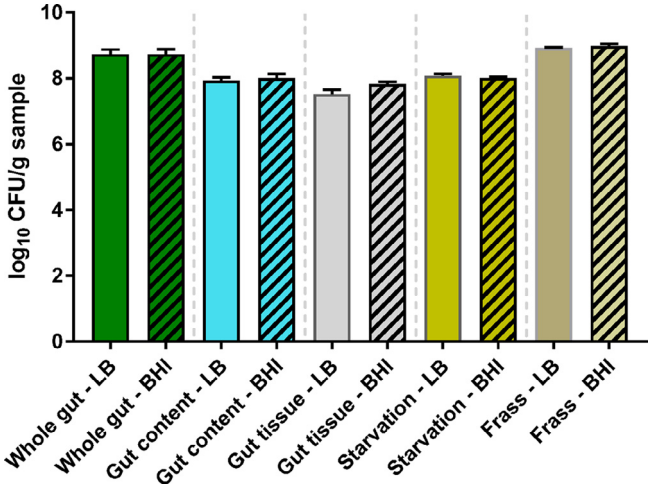

Fig. 4. Bacterial counts on nutrient agar indicating abundant bacteria in all five sample types. Each column represents six independent biological replicates. Error bars denote standard errors.

respectively), demonstrating the impact of sample types on microbial community structure in lepidopteran herbivores.

3.4. Variability in community composition among sample types

High-throughput 16S gene surveys determined the composition of the bacterial community across the five sample types under each DNA extraction condition, as shown in Fig. 6 (small-scale extraction) and 7 (large-scale extraction). All silkworms were from the same laboratory population reared under identical environmental conditions, and the overall interindividual variation in microbiota composition was low within each sample type (Figs. 6a, 7a). In addition, several OTUs were common to all five sample types, particularly with respect to the dominance of sequences in the phylum Firmicutes. *Enterococcus* was the most abundant genus, constituting up to 52.9% (small-scale extraction) and 35.9% (large-scale extraction) of all sequences obtained with each DNA extraction method. The genus *Acinetobacter*, belonging to the Proteobacteria phylum, was also found in most of our samples, making up to 20.6% (small-scale extraction) and 25.8% (large-scale extraction) of the total sequences (Figs. 6a, 7a). However, we observed significant differences in the distribution of various bacterial taxa depending on the sample type under both extraction conditions.

The small-scale, kit-based procedure showed that despite community representations being generally similar between whole gut and gut content sample types, the three other sample types tested (gut tissue, starvation and frass) were more distantly related, according to PERMANOVA pairwise analysis based on weighted UniFrac values (Fig. 6a, S2a). The whole gut samples captured the full composition of the microbial community. In addition to *Enterococcus* and *Acinetobacter*, the relative abundance of bacterial species, including *Enterobacter*, *Glutamicibacter*, *Rhodococcus*, *Corynebacterium* and *Sphingomonas*, was also high, and these bacteria were consistently observed in all individuals within this group. These OTUs were also found in gut content samples, which provided a community profile comparable to that of the whole gut samples. By contrast, other sample types, especially the frass samples, resulted in significant variation in taxonomic composition (gut tissue vs. whole gut, P = 0.0338; starvation vs. whole gut, P = 0.016; frass vs. whole gut, P = 0.0084, PERMANOVA test with 9999 permutations), and most taxa had a reduced proportional abundance (Figs. S2a, 6a). For the gut tissue, there was a relatively low abundance of *Enterococcus*, *Rhodococcus*, *Sphingomonas*, *Piscinibacter* and *Enterobacter*. Interestingly, we found that the abundance of *Glutamicibacter* and *Acinetobacter* was higher than that present in other sample types, suggesting that the two bacteria potentially attach to the gut epithelial tissue. When starved, the difference observed was that *Enterobacter* and *Glutamicibacter* abundance were greatly reduced, and *Enterococcus* abundance was increased.

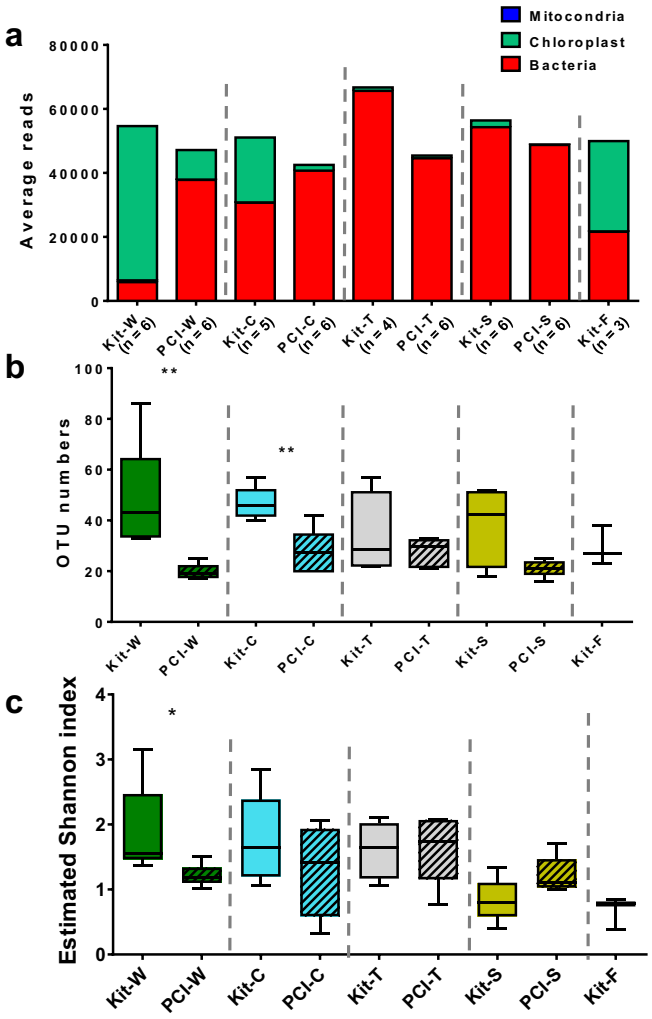

Fig. 5. Bacterial community diversity across sample types. (a) Average number of sequencing reads per sample type obtained with each DNA extraction method (Kit vs. PCI). (b) Changes in the number of OTUs. (c) Changes in the overall bacterial diversity. *, significance ($P < 0.05$, Student's *t* test); **, significance ($P < 0.01$, Student's *t* test). Kit and PCI represent the small-scale and large-scale metagenomic DNA extraction methods respectively. W, whole gut; C, gut content; T, gut tissue; S, starvation; F, frass. n, the number of biological replicates that were used for analyses.

In frass samples, compared to the whole gut, the microbiota was skewed towards only one dominant genus (*Enterococcus*, comprising as much as 88.3% of all sequences), resulting in a simple community composition that underestimated diversity (Fig. 5c).

We also evaluated the Spearman's distance (ranked species abundances) for each sample to assess the magnitude of microbiome variation between the samples as noted before [52]. The calculation of these distances produces a matrix where microbial taxa rather than samples were compared to one another. This Spearman's distance matrix represents the strength of the correlation among microbial pairs; thus, smaller distances represent stronger correlations, which were visualized using a color key (Fig. 6b). We found that the whole gut and gut content samples recovered comparable species rankings, while higher heterogeneity of the microbiota was evident in all other sample types, especially in the starvation and frass samples. PCoA analysis of pairwise weighted UniFrac values further revealed that most of the variation (>80%) was captured by the first two principal coordinates, and the clustering of sample types was easily observable (Fig. 6c). Notably, the frass and starvation samples showed larger overall dissimilarities in microbiota composition than did the other samples types.

Similarly, the large-scale PCI-based method also indicated highly variable patterns in bacterial community structure among sample types (Fig. 7a, S2b). Although good correlations between composition data from different biological replicates were observed in each sample type (Fig. 7b), bacterial communities were significantly separated among the four groups ($P < 0.01$, PERMANOVA test with 9999 permutations), which was evident in the PCoA diagram calculated with weighted UniFrac values (Fig. 7c). The whole gut showed a distinct microbial community from all other samples at the genus level with a high relative abundance of *Enterococcus* and *Acinetobacter*, and a lower abundance of other bacteria. *Ralstonia* was most abundant in the starvation samples, representing 41.3% of the total sequences. The same trend was also present in the small-scale, kit-based procedure (Fig. 6).

Collectively, biases in measures of community structure are consistent across different sample types, independent of the DNA extraction methodology that was used. The whole gut might be considered as the representative sample type for lepidopteran gut microbiome studies.

4. Discussion

An important prerequisite for the successful assessment of microbial communities is an efficient procedure for extracting DNA from representative samples [53]. Several previous studies have also revealed that multiple sources of contamination influence the characterization of bacterial communities in samples with low bacterial biomass [25,54]. We suggest using effective approaches (including decontam, negative controls) to identify potential contaminant sequences in a low-biomass environment too.

A direct examination of DNA extraction methods has never been described in the model Lepidoptera silkworm *B. mori* microbiome studies. Here, our comparative evaluations conducted using the same starting material clearly indicate that DNA extraction can also impact sequence-based silkworm microbiome analyses, as has recently been demonstrated for human fecal metagenomic studies [23–24]. In terms of DNA purity, there was no difference between the methods applied in this study. All DNA samples extracted could be used directly as a template for the amplification of fragments of the 16S rRNA gene under standard PCR conditions. The increased yield of DNA obtained by the small-scale, kit-based procedure compared with that derived from the large-scale, PCI-based procedure might be related to the more efficient extraction process of the former. Furthermore, the PCI method requires organic solvent and is time-consuming. Importantly, our survey of bacterial DNA extracted by these two procedures showed that the kit-based procedure was able to detect the underlying high-diversity community structure that was not realized with the PCI-based procedure. As demonstrated in control experiments, this enhanced detection seems to be achieved through efficient lysis of Gram-positive bacteria which possess a tough cell wall. The disruption/lysis of the bacterial envelopes and membranes can be expected to be biased for specific bacterial taxa due to differences in cell wall structure and integrity, contributing to variation in the reported abundance and diversity of gut bacteria. Similar to the results of the present work, Costea et al. highlighted the recovery of Gram-positive and Gram-negative bacteria as an important source of variation between extraction methods in human microbiome studies [23]. The caustic nature of the chemicals employed in the commercial extraction kit appeared to be able to lyse the bacterial cells more effectively. Altogether, we recommend a small-scale, kit-based methodology for surveys of microbial communities

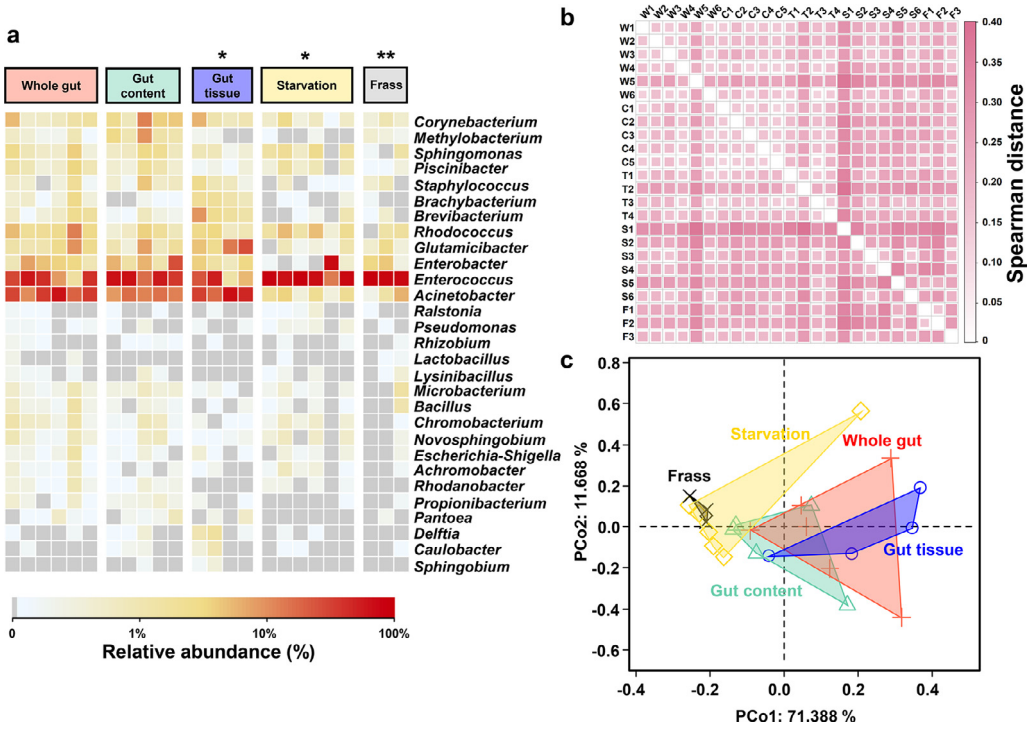

Fig. 6. Microbiota similarities and differences between sample types under small-scale extraction conditions. (a) Heat map showing the relative abundance of dominant taxa for each sample type. Each column represents an individual sample (biological replicate). *, P < 0.05; **, P < 0.01 (PERMANOVA test with 9999 permutations, see supplementary Fig. S2a). (b) Spearman distance matrix showing all the pairwise distances between samples, highlighting whether sample types may be considered comparable under measures of similarity, where a darker pink indicates larger variation. W, whole gut; C, gut content; T, gut tissue; S, starvation; F, frass. (c) PCoA analysis of pairwise weighted UniFrac values showing polygons that indicate clustering of bacterial compositions based on sample types, with geometrical shapes encompassing the range of each sample type. Each symbol represents a single silkworm individual (biological replicate), and colors indicate sample type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

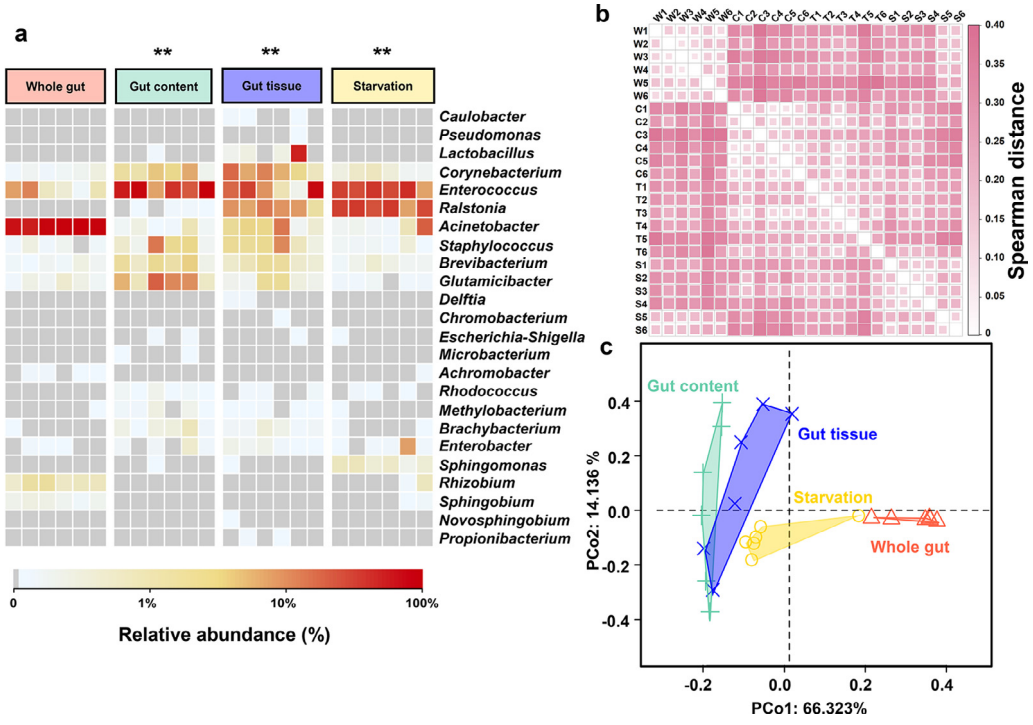

Fig. 7. Microbiota similarities and differences between sample types under large-scale extraction conditions. (a) Heat map showing the relative abundance of dominant taxa for each sample type. Each column represents an individual sample. **, P < 0.01 (PERMANOVA test with 9999 permutations, see supplementary Fig. S2b). (b) Spearman distance matrix shows all pairwise distances between samples, highlighting whether sample types may be considered comparable under measures of similarity, where a darker pink indicates larger variation. (c) PCoA analysis of pairwise weighted UniFrac values showing polygons that indicate clustering of bacterial compositions based on sample types, with geometrical shapes encompassing the range of each sample type. Each symbol represents a single silkworm individual (biological replicate), and colors indicate sample type. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

associated with lepidopteran insects due to the favorable trade-offs in automation, speed, and accuracy as well as the suitability for a highly parallel and high-throughput approach.

Not only the DNA extraction method but also the sample type has an extensive impact on gut microbiota analysis. Most studies of the gut microbiota of other insect models, such as the fruit fly, are based on DNA extracted from isolated whole-guts or whole-bodies, due to the relatively small sample sizes (typically at the milligram level per individual) [55]. In contrast, lepidopteran caterpillars consume large amounts of plant material, and body mass and volume increase dramatically in a short time (Fig. 1). Therefore, caterpillars are often starved, or their frass or gut tissues were used directly for DNA extraction. In the present study, we further addressed the effect of sample type (including the whole gut, gut content, gut tissue, starvation and frass) on the assessment of the gut microbiota of Lepidoptera. Our experiment demonstrated that all sample types resulted in adequate yields of microbial DNA, but community diversity and composition varied considerably among them.

In the case of the small-scale procedure, gut content and whole gut samples tended to be more comparable, and both sample types more fully captured the diversity of microbes (Fig. 6). Compared to the whole gut, the gut tissue and starvation samples displayed a significantly different microbiota as defined by the PERMANOVA test on weighted UniFrac data. Furthermore, even larger differences between community structures could be found in the frass samples, suggesting that the frass poorly represents the complexity of the gut microbiota. This is consistent with results published previously on the gut microbiotas of other animals, including humans [26–27,56]. Similarly, the large-scale procedure also demonstrated the important differences in the relative abundances of key bacterial members among sample types and highlighted that starvation, compared with the whole gut, results in extensive changes in bacterial diversity. The abundance of some taxa, including *Enterobacter*, decreased when caterpillars were starved, indicating transient bacteria associated with plant diet. However, starvation also stimulated a higher bacterial load, suggesting that some gut members multiply in the absence of food consumption. Consistent with these findings, studies in other animals, such as reindeer (*Rangifer tarandus*), have indicated that a period of starvation stresses the host and additionally affects the microbiome structure [57]. Based on these results, we also recommend sampling the whole gut under normal rearing conditions whenever possible, as this sample type provides the most accurate assessment of the gut microbiome.

A limitation of using the whole gut is the rather high percentage of chloroplasts. It is known that chloroplast sequence contamination of 16S rRNA gene sequencing analyses is particularly problematic when sampling bacterial communities in herbivorous insects [34]. We suggest including more biological replicates in the study design to retain the majority of species present in the herbivore gut microbiota. The filtration and density gradient centrifugation method for enrichment of the microbial community can also improve retention of bacterial diversity [41].

Enterococcus and *Acinetobacter* bacteria, which have been previously found to be the most predominant microorganisms within the silkworm gut microbiota [58], were also predominant in this study. Interestingly, *Enterococcus* sp. was found to be the dominant taxon in the laboratory-reared noctuid moth *Heliothis virescens*, but was completely absent from field *H. virescens* larvae, indicating possible shifts in microbial community profiles upon cultivation of the insect in the laboratory [59]. Previous studies also reported a great diversity of metabolic capabilities represented in the total gene pool of the gut microbiota in the silkworm [41,60]. The functions of the major gut colonizers deserve further study in this model organism.

5. Conclusions

Over 150,000 species of Lepidoptera have been identified so far, and this order presents highly diverse morphologies and behaviors and performs diverse functions in broad ecological niches, but only <0.1% have been screened for bacterial associates, which reveals that our knowledge of lepidopteran-associated bacteria is still very limited [61]. Our study highlights the importance of methodological consistency for accurate characterization and comparison of herbivore microbiological assemblages. We recommend a small-scale DNA extraction approach for sampling the whole gut under normal insect rearing conditions, which concurrently improved assay coverage and yielded favorable quantitative parameters for bacterial detection and quantification. Adoption of this methodology will minimize technical biases and facilitate a far more extensive microbiome study from a diverse array of lepidopteran species. It should be pointed out that non-destructively monitoring gut microbiota has some strengths when necessary. For instance, one could sample frass at multiple times and still rear caterpillars to adults to measure traits of interest (e.g., development time, survival rates, size) that could be influenced by/associated with gut microbes. Therefore, pros and cons of different sample types should be considered carefully in light of the research question(s) and constraints on experimental design.

With more standardization, control of sample processing and data analysis [62–63], increased concordance among different studies can be expected in the field, which might allow us to draw conclusions about the intimacy of the host–symbiont associations and to speculate on the possible coevolutionary relationships between lepidopteran herbivores and their symbiotic microbiota.

6. Data availability

Supporting information has been made available online. The datasets generated and analyzed during the current study are available in the NCBI SRA repository (Accession number SRP100894 and SRX3759742).

Author contributions

Y.S. developed the study concept and design. N.Z. and J.H. performed laboratory work. X.S. assisted with protocol development for metagenomic DNA extraction. C.S. participated in the sequencing, quality control and sequence alignment. J.H. analyzed all data. N.Z. performed the statistical analysis. Y.S. and A.M. were major contributors in writing the manuscript with input from others. All authors read and approved the final manuscript. The authors declare that they have no competing interests.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the editors from AJE for their editorial assistance. This work was supported by grants from the National Natural Science Foundation of China (Grant No. 32022081 and 31970483), China Agriculture Research System (Grant No. CARS-18-ZJ0302), key laboratory of silkworm and bee resource utilization and innovation of Zhejiang Province (2020E10025) and the Max Planck Society, Germany. The funders had no role in study design, data collection and interpretation, or writing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.08.020.

References

[1] Mithöfer A, Boland W. Plant defense against herbivores: chemical aspects. Annu Rev Plant Biol 2012;63(1):431–50.
[2] Karban R, Agrawal AA. Herbivore offense. Annu Rev Ecol Syst 2002;33 (1):641–64.
[3] Chen B, Du K, Sun C, Vimalanathan A, Liang X, Li Y, et al. Gut bacterial and fungal communities of the domesticated silkworm (Bombyx mori) and wild mulberry-feeding relatives. Isme J 2018;12(9):2252–62.
[4] Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc Natl Acad Sci U S A 2014;111(40):14500–5.
[5] Tobias NJ, Eberhard FE, Guarneri AA. Enzymatic biosynthesis of B-complex vitamins is supplied by diverse microbiota in the Rhodnius prolixus anterior midgut following Trypanosoma cruzi infection. Comput Struct Biotechnol J 2020;18:3395–401.
[6] Douglas AE. Multiorganismal insects: diversity and function of resident microorganisms. Annu Rev Entomol 2015;60(1):17–34.
[7] Mason CJ. Complex relationships at the intersection of insect gut microbiomes and plant defenses. J Chem Ecol 2020;46(8):793–807.
[8] Johnston PR, Rolff J, Schneider DS. Host and symbiont jointly control gut microbiota during complete metamorphosis. PLoS Pathog 2015;11(11): e1005246.
[9] Ignasiak K, Maxwell A. Oxytetracycline reduces the diversity of tetracycline-resistance genes in the Galleria mellonella gut microbiome. BMC Microbiol 2018;18:228.
[10] Allonsius CN, Van Beeck W, De Boeck I, Wittouck S, Lebeer S. The microbiome of the invertebrate model host Galleria mellonella is dominated by Enterococcus. Anim Microbiome 2019;1:7.
[11] Broderick NA, Raffa KF, Goodman RM, Handelsman Jo. Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. Appl Environ Microbiol 2004;70(1):293–300.
[12] Chen B, Teh B-S, Sun C, Hu S, Lu X, Boland W, et al. Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore Spodoptera littoralis. Sci Rep 2016;6(1). https://doi.org/10.1038/srep29505.
[13] Berman TS, Laviad-Shitrit S, Lalzar M, Halpern M, Inbar M. Cascading effects on bacterial communities: cattle grazing causes a shift in the microbiome of a herbivorous caterpillar. ISME J 2018;12(8):1952–63.
[14] Vilanova C, Baixeras J, Latorre A, Porcar M. The generalist inside the specialist: gut bacterial communities of two insect species feeding on toxic plants are dominated by Enterococcus sp.. Front Microbiol 2016;7:1005.
[15] Brinkmann N, Martens R, Tebbe CC. Origin and diversity of metabolically active gut bacteria from laboratory-bred larvae of Manduca sexta (Sphingidae, Lepidoptera, Insecta). Appl Environ Microbiol 2008;74(23):7189–96.
[16] Xia X, Gurr GM, Vasseur L, Zheng D, Zhong H, Qin B, et al. Metagenomic sequencing of diamondback moth gut microbiome unveils key holobiont adaptations for herbivory. Front Microbiol 2017;8. https://doi.org/10.3389/fmicb.2017.00663.
[17] Montagna M, Mereghetti V, Gargari G, Guglielmetti S, Faoro F, Lozzia G, et al. Evidence of a bacterial core in the stored products pest Plodia interpunctella: the influence of different diets. Environ Microbiol 2016;18(12):4961–73.
[18] Mereghetti V, Chouaia B, Limonta L, Locatelli DP, Montagna M. Evidence for a conserved microbiota across the different developmental stages of Plodia interpunctella. Insect Sci 2019;26(3):466–78.
[19] Russell JA, Funaro CF, Giraldo YM, Goldman-Huertas B, Suh D, Kronauer DJC, et al. A veritable menagerie of heritable bacteria from ants, butterflies, and beyond: broad molecular surveys and a systematic review. PLoS ONE 2012;7 (12):e51027.
[20] Kaiser W, Huguet E, Casas J, Commin C, Giron D. Plant green-island phenotype induced by leaf-miners is mediated by bacterial symbionts. Proc Biol Sci/The Royal Society 2010;277(1692):2311–9.
[21] Hammer TJ, Janzen DH, Hallwachs W, Jaffe SP, Fierer N. Caterpillars lack a resident gut microbiome. Proc Natl Acad Sci U S A 2017;114(36):9641–6.
[22] Ravenscraft A, Berry M, Hammer T, Peay K, Boggs C. Structure and function of the bacterial and fungal gut microbiota of Neotropical butterflies. Ecol Monogr 2019;89(2):e01346. https://doi.org/10.1002/ecm.2019.89.issue-210.1002/ecm.1346.
[23] Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017;35(11):1069–76.
[24] Zmora N et al. Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. Cell 2018;174:1388–1405 e1321.
[25] Paniagua Voirol LR, Valsamakis G, Yu M, Johnston PR, Hilker M. How the 'kitome' influences the characterization of bacterial communities in lepidopteran samples with low bacterial biomass. J Appl Microbiol 2021;130 (6):1780–93.
[26] Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, et al. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J 2014;8(4):881–93.
[27] Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans ADL, de Vos WM. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol 2002;68(7):3401–7.
[28] Liu H, Chen BS, Hu SR, Liang XL, Lu XM, Shao YQ. Quantitative proteomic analysis of germination of nosema bombycis spores under extremely alkaline conditions. Front Microbiol 2016;7:1459.
[29] Martinez-Solis M, Collado MC, Herrero S. Influence of diet, sex, and viral infections on the gut microbiota composition of Spodoptera exigua caterpillars. Front Microbiol 2020;11:753.
[30] Deguenon JM, Dhammi A, Ponnusamy L, Travanty NV, Cave G, Lawrie R, et al. Bacterial microbiota of field-collected Helicoverpa zea (Lepidoptera: Noctuidae) from transgenic Bt and Non-Bt cotton. Microorganisms 2021;9 (4):878.
[31] Grunert LW, Clarke JW, Ahuja C, Eswaran H, Nijhout HF, Doucet D. A Quantitative analysis of growth and size regulation in Manduca sexta: the physiological basis of variation in size and age at metamorphosis. PLoS ONE 2015;10(5):e0127988.
[32] Sudakaran S, Retz F, Kikuchi Y, Kost C, Kaltenpoth M. Evolutionary transition in symbiotic syndromes enabled diversification of phytophagous insects on an imbalanced diet. Isme J 2015;9(12):2587–604.
[33] Kaltenpoth M, Winter SA, Kleinhammer A. Localization and transmission route of Coriobacterium glomerans, the endosymbiont of pyrrhocorid bugs. FEMS Microbiol Ecol 69, 373-383 (2009).
[34] Hanshew AS, Mason CJ, Raffa KF, Currie CR. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. J Microbiol Methods 2013;95(2):149–55.
[35] Maltz MA, Bomar L, Lapierre P, Morrison HG, McClure EA, Sogin ML, et al. Metagenomic analysis of the medicinal leech gut microbiota. Front Microbiol 2014;5. https://doi.org/10.3389/fmicb.2014.00151.
[36] Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM. The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isoamyl alcohol DNA extraction. Mol Ecol Resour 2015;15:168–76.
[37] Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, et al. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombyx). Science 2009;326(5951):433–6.
[38] Liang X, Sun C, Chen B, Du K, Yu T, Luang-In V, et al. Insect symbionts as valuable grist for the biotechnological mill: an alkaliphilic silkworm gut bacterium for efficient lactic acid production. Appl Microbiol Biotechnol 2018;102(11):4951–62.
[39] Shao Y, Arias-Cordero EM, Boland W. Identification of metabolically active bacteria in the gut of the generalist Spodoptera littoralis via DNA stable isotope probing using 13C-glucose. J Visual Exp JoVE 2013;81:e50734.
[40] Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, et al. Topographic diversity of fungal and bacterial communities in human skin. Nature 2013;498 (7454):367–70.
[41] Chen B, Yu T, Xie S, Du K, Liang X, Lan Y, et al. Comparative shotgun metagenomic data of the silkworm Bombyx mori gut microbiome. Sci Data 2018;5(1):180285.
[42] Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 2011;27(21):2957–63.
[43] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884-i890 (2018).
[44] Song Y, Jiang C-Y, Liang Z-L, Wang B-J, Jiang Y, Yin Ye, et al. Casimicrobium huifangae gen. nov., sp. nov., a Ubiquitous "Most-Wanted" Core bacterial taxon from municipal wastewater treatment plants. Appl Environ Microbiol 2020;86(4). https://doi.org/10.1128/AEM.02209-19.
[45] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460-2461 (2010).
[46] Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35(21):7188–96.
[47] Chen X, Krug L, Yang H, Li H, Yang M, Berg G, et al. Nicotiana tabacum seed endophytic communities share a common core structure and genotype-specific signatures in diverging cultivars. Comput Struct Biotechnol J 2020;18:287–95.
[48] Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018;6:226.
[49] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75(23):7537–41.
[50] Willis AD, Martin BD. Estimating diversity in networked ecological communities. Biostatistics 2020.
[51] Tang Z-Z, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics 2016;32(17):2618–25.
[52] Williams RJ, Howe A, Hofmockel KS. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Front Microbiol 2014;5:358.

[53] Hermans SM, Buckley HL, Lear G. Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. Mol Ecol Resour 2018;18(3):557–69.
[54] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. Bmc Biol 2014;12(1). https://doi.org/10.1186/s12915-014-0087-z.
[55] Martinson VG, Douglas AE, Jaenike J, Coulson T. Community structure of the gut microbiota in sympatric species of wild Drosophila. Ecol Lett 2017;20(5):629–39.
[56] Hayashi H, Takahashi R, Nishi T, Sakamoto M, Benno Y. Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. J Med Microbiol 2005;54:1093–101.
[57] Præsteng KE, Pope PB, Cann IKO, Mackie RI, Mathiesen SD, Folkow LP, et al. Probiotic dosing of Ruminococcus flavefaciens affects rumen microbiome structure and function in reindeer. Microb Ecol 2013;66(4):840–9.
[58] Sun Z, Lu Y, Zhang H, Kumar D, Liu B, Gong Y, et al. Effects of BmCPV infection on silkworm Bombyx mori intestinal bacteria. PLoS ONE 2016;11(1):e0146313.
[59] Staudacher H, Kaltenpoth M, Breeuwer JAJ, Menken SBJ, Heckel DG, Groot AT, et al. Variability of bacterial communities in the moth Heliothis virescens indicates transient association with the host. PLoS ONE 2016;11(5):e0154514.
[60] Chen B, Zhang N, Xie S, Zhang X, He J, Muhammad A, et al. Gut bacteria of the silkworm Bombyx mori facilitate host resistance against the toxic effects of organophosphate insecticides. Environ Int 2020;143:105886.
[61] Paniagua Voirol LR, Frago E, Kaltenpoth M, Hilker M, Fatouros NE. Bacterial symbionts in lepidoptera: their diversity, transmission, and impact on the host. Front Microbiol 2018;9:556.
[62] Aguirre de Cárcer D. Experimental and computational approaches to unravel microbial community assembly. Comput Struct Biotechnol J 2020;18:4071–81.
[63] Bokulich NA, Ziemski M, Robeson MS, Kaehler BD. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Comput Struct Biotechnol J 2020;18:4048–62.