


## ORIGINAL ARTICLE

# How the 'kitome' influences the characterization of bacterial communities in lepidopteran samples with low bacterial biomass

L.R. Paniagua Voirol<sup>1</sup>, G. Valsamakis<sup>1</sup>, M. Yu<sup>1</sup>, P.R. Johnston<sup>2,3,4</sup> and M. Hilker<sup>1</sup> <sup>1</sup> Applied Zoology/Animal Ecology, Institute of Biology, Freie Universität Berlin, Berlin, Germany<sup>2</sup> Evolutionary Biology, Institute of Biology, Freie Universität Berlin, Berlin, Germany<sup>3</sup> Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Berlin, Germany<sup>4</sup> Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany**Keywords**

bacterial community structure, diversity, DNA extraction, insect-gut microbiome, kitome.

**Correspondence**

Monika Hilker, Applied Zoology/Animal Ecology, Institute of Biology, Freie Universität Berlin, Haderslebener Str. 9, 12163, Berlin, Germany.

E-mail: [monika.hilker@fu-berlin.de](mailto:monika.hilker@fu-berlin.de)

2020/1469: received 16 July 2020, revised 2 October 2020 and accepted 23 October 2020

doi:10.1111/jam.14919

**Abstract****Aims:** We aimed to elucidate whether the DNA extraction kit and bacteria therein affect the characterization of bacterial communities associated with butterfly samples harbouring different bacterial abundancies.**Methods and Results:** We analysed bacteria associated with eggs of *Pieris brassicae* and with adults of this butterfly, which were either untreated or treated with antibiotics (ABs). Three DNA extraction kits were used. Regardless of the extraction kit used, PCR amplification of the bacterial 16S rRNA gene detected very low bacterial presence in eggs and AB-treated butterflies. In untreated butterflies, bacterial signal intensity varied according to the kit and primers used. Sequencing (MiSeq) of the bacterial communities in untreated and AB-treated butterflies revealed a low alpha diversity in untreated butterflies because of the dominance of few bacteria genera, which were detectable regardless of the kit. However, a significantly greater alpha diversity was found in AB-treated butterflies, evidencing a true bias of the results due to bacterial contaminants in the kit.**Conclusions:** The so-called 'kitome' can impact the profiling of Lepidoptera-associated bacteria in samples with low bacterial biomass.**Significance and Impact of the Study:** Our study highlights the necessity of method testing and analysis of negative controls when investigating Lepidoptera-associated bacterial communities.**Introduction**

The molecular analysis of bacterial communities has greatly extended our understanding of these hidden members of ecosystems (e.g. Torsvik and Ovreas 2002; Riesenfeld *et al.* 2004; Bringel and Couee 2015; Lievens *et al.* 2015). Valid comparisons of bacterial communities and analyses of their successions require precise and reproducible descriptions of their composition. The results of the analysis of bacterial communities by PCR and sequencing techniques sensitively depend on a wide range of parameters (Wintzingerode *et al.* 1997; Fouhy *et al.* 2016). For instance, storage temperatures of samples and preservation buffers can affect the detected bacterial community diversity, richness and relative abundance

(Choo *et al.* 2015). DNA extraction methods using different cell lysis procedures have an impact on absolute microbial numbers, community richness and relative abundance (Ariefdjohan *et al.* 2010; Henderson *et al.* 2013). Furthermore, the choice of primers targeting the bacterial 16S rRNA gene can significantly affect which members are detected. The so-called 'universal primers' (e.g. Ben-Dov *et al.* 2006), in fact, vary in their efficacy to cover the richness of bacteria present in a sample (Baker *et al.* 2003). In addition, the different 16S rRNA gene regions amplified by different primers produce varying results when analysing community diversity by next-generation sequencing (Bukin *et al.* 2019).

Contaminating bacterial DNA is commonly found in different DNA extraction kits (Salter *et al.* 2014). This so-

called 'kitome' can have a great impact on samples with a low bacterial abundance. Thus, when using DNA sequencing-based techniques, negative controls should always be used in parallel to identify those bacterial members belonging to the actual sample (Wintzingerode *et al.* 1997; Salter *et al.* 2014). A wide range of studies of bacterial communities, especially in soil, water, vertebrate animals and humans have shown that the treatment of samples prior to analysis, the DNA extraction method, the primers, the sequencing platform and the purity of reagents greatly affect the description of these bacterial communities (Methé *et al.* 1998; Martin-Laurent *et al.* 2001; Cuív *et al.* 2011; Gilbert *et al.* 2012; Henderson *et al.* 2013; Burbach *et al.* 2016; Castelino *et al.* 2017).

Bacterial communities associated with insects have been recognized as important players that can greatly affect their host biology and its interaction with other trophic levels (Feldhaar 2011; Ferrari and Vavre 2011; Hansen and Moran 2014; Douglas 2015; Paniagua Voirol *et al.* 2018). However, a great number of economically important insect pests remain unexplored in terms of their microbiota, although knowledge of their microbiota might contribute to a deeper understanding of, for example, development of insect resistance against insecticides.

Lepidoptera (moths and butterflies) are a taxon comprising numerous agricultural and forestry pest species. The extent to which bacterial associates affect the biology of Lepidoptera is difficult to assess. The studies characterizing the bacterial communities associated with various moths and butterflies are hardly comparable because they employ different DNA extraction techniques and 16S primers (Paniagua Voirol *et al.* 2018). Furthermore, studies of the lepidopteran microbiota are mostly focused on the plant-damaging larval stage (Paniagua Voirol *et al.* 2018), and analyses on other life stages, such as the eggs or adults, are scarce (see: Chen *et al.* 2016; Phalnikar *et al.* 2018; Ravenscraft *et al.* 2019), although bacterial associates in these stages may be relevant for the insect's fitness. In fact, it has been found that bacterial abundance in lepidopteran adults can be much higher than in other life stages (Hammer *et al.* 2014, 2017; Ravenscraft *et al.* 2019; Paniagua Voirol *et al.* 2020).

Treatments with antibiotics (ABs) are commonly used to manipulate bacterial communities associated with Lepidoptera (Paniagua Voirol *et al.* 2018, 2020). Samples from AB-treated individuals are frequently analysed along with untreated control samples to assess differences in bacterial community composition. However, no study addressed so far the question whether different DNA extraction kits differentially influence the characterization of bacterial communities in Lepidoptera samples with different bacterial abundance. The aim of this study is to find out whether the choice of DNA extraction kit is

critical when investigating bacterial communities associated with Lepidoptera samples, which contain bacteria in different abundancies.

We studied the bacterial communities associated with *P. brassicae* eggs and adults harbouring different bacterial loads. Eggs were expected to have a low bacterial load (see Paniagua Voirol *et al.* 2020). To produce samples of *P. brassicae* adults with different bacterial load, butterflies were treated with ABs. A significant reduction of the bacterial abundance in *P. brassicae* butterflies upon AB treatment was previously confirmed *via* qPCR and reported by Paniagua Voirol *et al.* (2020).

We processed the samples using three different commercially available DNA extraction kits and addressed the following questions: (i) Does the applied kit, in combination with different 16S primers, affect the efficacy of the PCR amplification of the 16S rRNA gene from bacteria in eggs, AB-treated and untreated adults? (ii) How does the kit affect the detection of bacterial taxa in untreated and AB-treated adults? We studied this question by sequencing (MiSeq) the bacterial communities and identified bacteria that were consistently present in adults as well as bacteria associated with the kits used. (iii) Therefore, we asked to which extent does this so-called 'kitome' shape the detected bacterial community in untreated and AB-treated butterflies; here, we compared the bacterial communities associated with butterflies with those obtained from negative extraction controls (NECs). We determined the dissimilarities and alpha diversities of the bacterial communities of the differently processed butterfly samples and of the NECs.

Our study shows that the results of the PCR amplification of the 16S rRNA gene vary in dependence of the kit used. The sequencing analysis revealed that the bacterial taxa detected in samples of control *P. brassicae* with high bacteria abundance were similar, regardless of the type of kit used. However, sequencing of samples with a low bacterial load and NECs evidenced bacterial contamination coming from the DNA extraction kits, which greatly shaped the detected bacterial communities and alpha diversities in these samples. Our study exposes that a valid characterization of the bacterial community associated with Lepidoptera, especially with those harbouring a low bacterial load, needs method testing and obligatory consideration of NECs.

## Materials and Methods

### Insect rearing

Insects originated from a laboratory-reared colony (Institute of Biology, Applied Zoology/Animal Ecology, Freie Universität Berlin, Germany). Larvae of the Large White

(*P. brassicae*) were reared on Brussels sprouts plants (*Brassica oleracea* var. *gemmifera*) in a climate chamber (18-h/6-h light/dark cycle,  $160 \mu\text{mol m}^{-2} \text{s}^{-1}$  light intensity,  $20^\circ\text{C}$  and r.h. 70%) until pupation. The plants were grown in a greenhouse. Pupae were transferred to a separate climate chamber (18-h/6-h light/dark cycle,  $220 \mu\text{mol m}^{-2} \text{s}^{-1}$  light intensity,  $23^\circ\text{C}$  and r.h. 70%), where adult butterflies emerged. Adults were fed with a 15% w/v honey solution provided in 1.5-ml Eppendorf tubes placed in the centre of artificial flowers. This insect line is here referred to as untreated or control line.

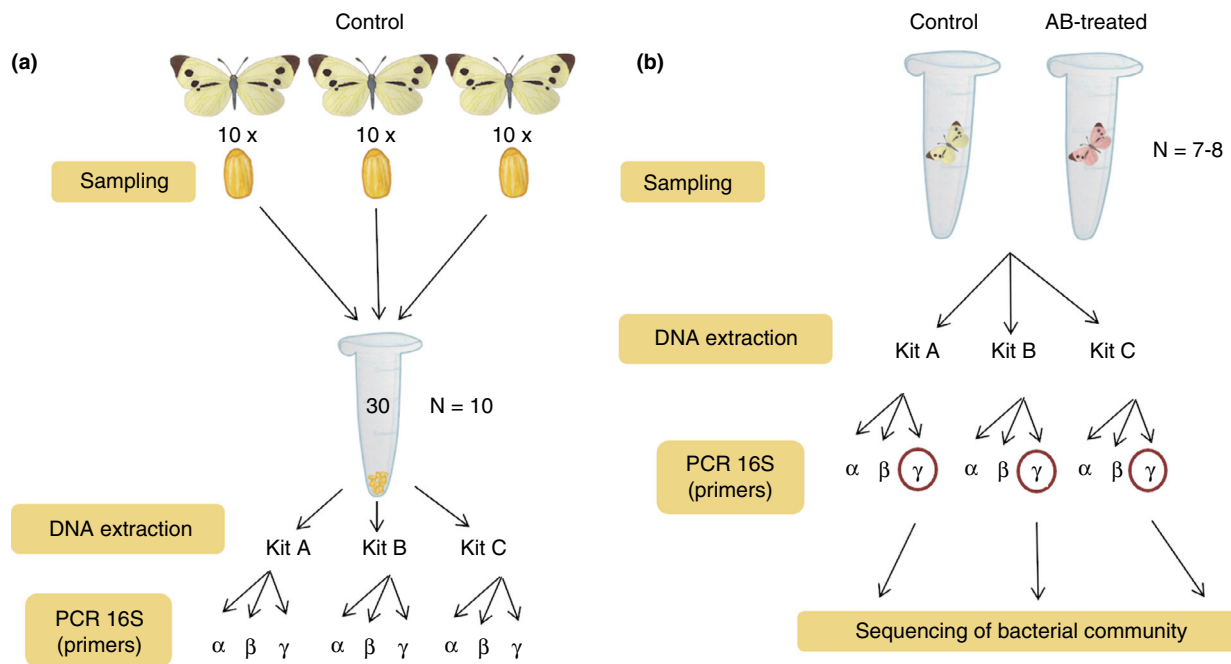
To elucidate how the choice of DNA extraction kit and bacterial contaminants therein affect the results when analysing insects with low bacterial abundance, we established an AB-treated line of *P. brassicae*. This AB-treated line was derived from the original *P. brassicae* rearing. For this line, larvae were reared on 7-week-old Brussels sprouts plants sprayed with a cocktail of four ABs (ampicillin, chloramphenicol, rifampicin and streptomycin), each in a concentration of  $0.5 \text{ mg ml}^{-1} \text{ H}_2\text{O}$  until the plant surface was uniformly covered with the solution. After pupation, the treatment with ABs was continued by feeding the butterflies with an AB-spiked aqueous honey

solution (15% w/v honey dissolved in the above-mentioned mixture of ABs).

### Sampling procedures

An egg sample (biological replicate) of *P. brassicae* consisted of a pool of 30 eggs collected from three different female butterflies (i.e. 10 eggs per female). We pooled eggs from three different females to minimize variation originating from individual females. For the collection of *P. brassicae* eggs, females laid egg clutches on the sterile side of a petri dish, while the other side was covered with a Brussels sprouts leaf to stimulate oviposition. The freshly laid eggs were further processed for DNA extraction (see below). In total, we analysed the bacterial abundance in 10 egg samples (biological replicates) by PCR (Fig. 1a).

We sampled adult control and AB-treated butterflies 2 weeks after emerging from the pupal stage. Each biological replicate consisted of an individual female after the removal of its wings. A total of 10 control and 10 AB-treated females were collected and processed with the different DNA extraction kits and subsequent PCR



**Figure 1** Experimental scheme of the sampling and further processing of *Pieris brassicae* samples. (a) Eggs were pooled from three females per sample to limit variation of bacteria detection across females. One egg sample consisted of 30 eggs obtained from three females. (b) Each butterfly sample consisted of a single female that was either untreated (control) or treated with antibiotics (AB). Samples were homogenized and TE buffer ( $280 \mu\text{l}$ ) was added to split the homogenate in three aliquots of  $50 \mu\text{l}$  each. Aliquots from the same sample were processed with each DNA extraction kit (A, B and C) (compare Table 1). The sample obtained by each DNA extraction was used for PCR amplification of the 16S rRNA gene using three different pairs of primers ( $\alpha$ ,  $\beta$  and  $\gamma$ ). For the adults, the product yielded with  $\gamma$  primers (encircled) was then used for bacterial community sequencing due to its clear signal and adequate amplicon size.  $N$  represents the number of biological replicates (biological samples).

amplification of the 16S rRNA gene. For the sequencing of the bacterial community, we processed seven to eight adult individuals of each the AB-treated and control butterflies (Fig. 1b).

### DNA extraction

Samples of *P. brassicae* eggs and adults were frozen in liquid nitrogen immediately after collection in sterile FastPrep® tubes and bead-ground for 15 s at 4500 rev min<sup>-1</sup> in a tissue homogenizer (Precellys Evolution®).

We used three different DNA extraction kits (Table 1). The kits were chosen based on whether they have been reported for successful DNA extraction of bacteria (i) in insect eggs (Pankewitz *et al.* 2007), (ii) in insect larvae or adults (Kaltenpoth *et al.* 2011; Salem *et al.* 2013; Leonhardt and Kaltenpoth 2014; Dohet *et al.* 2016; Engl *et al.* 2016; Staudacher *et al.* 2016; Flórez *et al.* 2017) or (iii) in a diverse range of samples (e.g. Lindh and Lehane 2011; Koga and Moran 2014; Dalla-Costa *et al.* 2017; Pendleton *et al.* 2017; Evans *et al.* 2018).

To test each biological replicate with the three different DNA extraction kits, 280 µl of TE buffer was added to the homogenized samples, followed by brief vortexing. Then, three aliquots of 50 µl were taken from the homogenate and processed with each kit. We added the corresponding lysis buffer from each kit to the aliquots of the homogenized samples from adults and eggs, thereby following the manufacturer's instructions. The DNA extraction was continued as indicated in the manufacturer's instruction. Each sample was subjected to a lysozyme (Epicentre Ready-Lyse™) digestion for a period of 30 min following the Epicentre Ready-Lyse™ protocol to maximize the lysis of Gram-positive bacteria (Ketchum *et al.* 2018). Samples were further processed exactly as indicated by each kit's protocol.

We included NECs, consisting of mock samples (containing no insect sample), and positive controls,

consisting of samples containing a resuspended (50 µl TE buffer) pellet of *Escherichia coli* DH5-α previously cultured in 1 ml of Luria Bertani (LB) broth at 37°C overnight. Extraction controls were processed at the same time as the true samples to control for kit-associated contamination and effectiveness of the bacterial DNA extraction.

Since contamination during DNA extraction can originate from different sources other than the kit itself (McFeters *et al.* 1993; McAlister *et al.* 2002; Witt *et al.* 2009; Motley *et al.* 2014), extractions were performed under clean bench conditions. All samples processed with each kit were handled at once including their respective controls. Plastic consumables and additional reagents (isopropanol, ethanol and molecular grade water) were of the same lot for all extractions.

Following the extraction, DNA concentrations were measured using the Thermo Scientific™ µDrop™ Plate and applying the manufacturer's instructions. The A<sub>260</sub>/A<sub>280</sub> and A<sub>260</sub>/A<sub>230</sub> ratios were verified to be within the values known for high-quality DNA samples (Lucena-Aguilar *et al.* 2016).

### Primers and PCR amplification of the 16S rRNA gene

For the PCR amplification of the bacterial 16S rRNA gene in the samples extracted with the types of kits mentioned above, we used three different pairs of primers (α, β and γ) commonly used for detection of bacteria *via* PCR (Table 2; Unno 2015; Eldridge *et al.* 2017; Thompson *et al.* 2017).

Primers targeting the V3–V4 regions of the 16S rRNA gene are considered more efficient at capturing bacterial community composition when compared to primers targeting other regions (e.g. V1–V3 region) (Castelino *et al.* 2017). Furthermore, the V3–V4 region has been targeted in important studies such as the Human Microbiome Project (Huttenhower *et al.* 2012; Fadrosch *et al.* 2014).

**Table 1** Overview of applied DNA extraction kits (A, B and C) and kit-specific information

Kit code	Commercial name	Extraction principle	Reason for testing	Specifications/Remarks
A	GE Healthcare illustra™ Tissue and Cells genomicPrep Mini Spin Kit	Column-based	Used to detect <i>Wolbachia</i> bacteria in insect eggs*	Protocol for extraction of genomic DNA from animal tissues
B	Epicentre Master Pure™ Purification Kit	Solution-based	Used in several studies on insect bacterial symbionts†	Protocol for tissue samples
C	Qiagen DNeasy® Blood and Tissue Kit	Column-based	Used in over 1000 publications for different purposes	Purification of total DNA from animal tissues (Spin-column protocol)

\*Pankewitz *et al.* (2007), the manufacturer of the kit A used by Pankewitz *et al.* changed, but the kit stayed the same.

†Dohet *et al.* (2016), Engl *et al.* (2016), Flórez *et al.* (2017), Kaltenpoth *et al.* (2011), Leonhardt and Kaltenpoth (2014), Salem *et al.* (2013) and Staudacher *et al.* (2016).

**Table 2** List of the three pairs of primers ( $\alpha$ ,  $\beta$  and  $\gamma$ ) used for PCR amplification of the 16S rRNA gene

Primer code	Name	Approx. amplicon size (bp)	Sequence (5'> 3')
$\alpha$	515F	290	GTGYCAGCMGCCGCGGTAA
	806R		GGACTACNVGGGTWTCTAAT
$\beta$	fD1	1500	AGAGTTTGATCCTGGCTCAG
	rP2		ACGGCTACCTTGTACGACTT
$\gamma^*$	341F	460	CCTACGGGNGGCWGCAG
	805R		GACTACHVGGGTATCTAATCC

\*Primers with an overhang for sequencing (5'> 3'): TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG (341F), GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (805R).

Therefore, we used two pairs of primers ( $\alpha$  and  $\gamma$ ) targeting such region. Primers  $\alpha$  are recommended by the Earth Microbiome Project (Thompson *et al.* 2017), whereas primers  $\gamma$  are specifically recommended by Illumina for the MiSeq platform (Illumina 2013; Klindworth *et al.* 2013). Although no NGS platform is available that allows multiple-sample sequencing of the complete 16S rRNA gene, we also included a pair of primers ( $\beta$ ), which amplify almost the entire 16S rRNA gene (Weisburg *et al.* 1991; Andreolli *et al.* 2013). Thus, we could assess whether the bacteria signal *per se* was influenced by the DNA extraction kit.

The PCR was conducted using the JumpStart™ Taq ReadyMix™ from Sigma Aldrich with 50 ng of DNA template in a total volume of 50  $\mu$ l. The PCR mixture was prepared under clean bench conditions. All consumables and reagents were of the same batch. The cycling parameters were as follows: an initial denaturation cycle at 94°C for 2 min followed by 30 cycles of denaturation at 94°C during 30 s, annealing (primers  $\alpha$ : 52°C,  $\beta$ : 55°C,  $\gamma$ : 55°C) for 30 s, extension at 72°C for 2 min and a final extension cycle at 72°C for 5 min.

In addition to the positive and NECs, we included positive and negative PCR controls to validate the effectiveness of the reaction. Positive PCR controls consisted of PCR mixtures containing 50 ng of *E. coli* DH5- $\alpha$  DNA as template. Negative PCR controls consisted of only the PCR mixtures containing no true samples. A volume of 10  $\mu$ l of the PCR product was run on a 1.2% agarose gel stained with ethidium bromide in 1X TAE buffer at 150 V for 30 min. Bands were visualized under UV light. The rest of the PCR product was kept for downstream sequencing.

### Bacterial community sequencing

For bacterial community sequencing, we selected the PCR products given by the  $\gamma$  primers, which targeted bacterial 16S sequences. These primers reliably produced a single

conspicuous band (amplicon) with a suitable size (<500 bp) for the sequencing platform. We sequenced (Illumina MiSeq platform) amplicons from butterflies (control and AB-treated) processed with the three different DNA extraction kits and the respective NECs, to assess the extent to which the 'kitome' shapes the detected bacterial community in samples with high and low bacterial abundance. Amplicons from eggs were not further sequenced as the bacterial load in AB-treated butterflies and eggs has been reported to be similarly low (see Paniagua Voirol *et al.* 2020).

PCR products were purified with AMPure beads (Beckmann Coulter, Brea, CA) and ligated to barcoded Illumina adapters *via* PCR using a high-fidelity DNA polymerase (Herculase II Fusion DNA Polymerase, Agilent). The cycling parameters were as follows: an initial denaturation cycle at 95°C for 2 min followed by eight cycles of denaturation at 95°C during 20 s, annealing at 52°C for 30 s, extension at 72°C for 30 s and a final extension cycle at 72°C for 3 min. Each library contained a specific combination of index adapters (dual-indexed) to allow later discrimination of samples after pooling. Concentrations of the ligated products were measured using Qubit® fluorometer (Thermo Fisher Scientific) and brought to an equimolar ratio prior to pooling. The pooled, barcoded samples were sequenced at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv) on the Illumina MiSeq platform (Illumina, San Diego, CA) using v3 600 cycles of (paired-end) sequencing. Sequencing reads were trimmed, denoised and overlapped using a full-stack R (R core Team 2018) pipeline incorporating dada2 (Callahan *et al.* 2016a, 2016b) and phyloseq (McMurdie and Holmes 2013). Forward and reverse reads were trimmed to 275 and 175 bp, respectively, truncated at the first instance of a quality score less than 2 and filtered on a maximum expected error rate of two errors per truncated read. The remaining forward and reverse reads were dereplicated and denoised using a parameterized model of substitution errors (Callahan *et al.* 2016a, 2016b). The resulting denoised read pairs were merged and subjected to *de novo* chimera removal. Taxonomy was assigned using the Ribosomal Database Project (Cole *et al.* 2014) training set, version 16. The resulting exact sequence variants (see Table S1 for sequencing variant numbers per sample) were agglomerated at the genus level. The R script applied for the sequencing analysis is provided in File S1.

### Statistical Analysis

Shannon indices were calculated using absolute counts (McMurdie and Holmes 2013) and analysed with Wald tests of linear models with linear combinations of

parameters using the contrast package (Kuhn *et al.* 2013). We did not rarefy or normalize the libraries for alpha diversity determination because library sizes were not significantly different among the treatment groups (Willis 2019) (see Table S2). Chloroplast sequences were present in small amounts in most samples (see Table S3) and filtered out for the analysis. Plant or insect mitochondrial sequences were present in very minor, negligible amounts (Table S3). Bray–Curtis distance matrices were calculated using relative abundance data and principal coordinate analysis (PCoA) ordination was performed using phyloseq (McMurdie and Holmes 2013). Multivariate analysis of variance was tested using vegan (Oksanen *et al.* 2018). Differential abundance testing was performed using DESeq2 (Love *et al.* 2014) in conjunction with phyloseq. Briefly, genus-level counts were modelled using generalized linear models of the negative binomial family with a logarithmic link. For all analyses, dispersion parameters were estimated with a local fit and empirical Bayes shrinkage. Specifically, likelihood-ratio tests were performed to test for the main effect of the applied kit and for a kit by treatment interaction. Counts for each genus were normalized by size factors accounting for variation in sequencing depth across samples. Size factors were estimated using the ‘median of ratios’ method described by equation 5 in Anders and Huber (2010). A modified geometric mean was used by taking the  $n$ th root of the product of the non-zero counts (McMurdie and Holmes 2013). The impact of multiple testing correction was mitigated by independent filtering using the mean normalized count for each genus across all samples. Genera were considered to be differentially abundant at FDR-corrected  $P < 0.05$ .

## Results

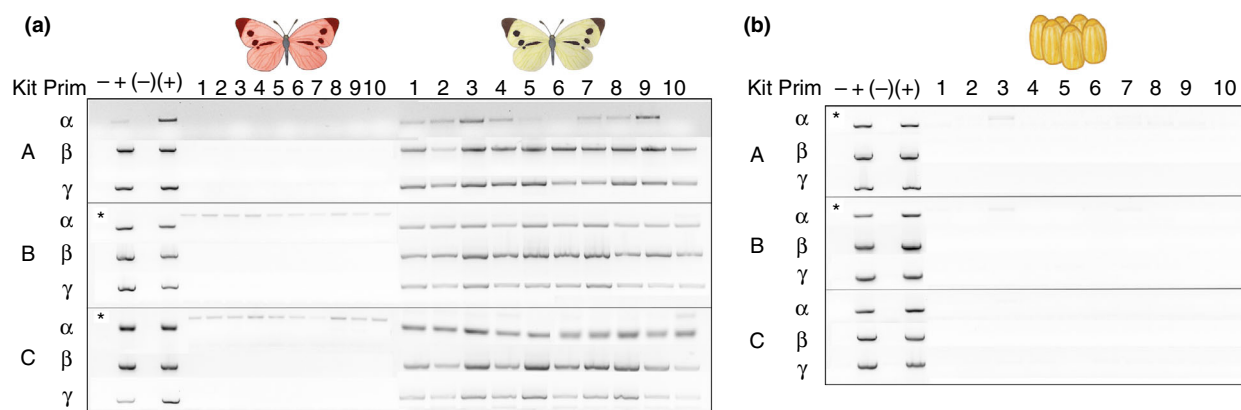
### Effects of the DNA extraction kit on PCR amplification of the 16S rRNA gene

We extracted DNA from *P. brassicae* eggs, AB-treated and untreated (control) adults using three different DNA extraction kits. We assessed whether PCR amplification of the bacterial 16S rRNA gene varies in dependence of the kit used to process the sample.

For control *P. brassicae* adults, the intensity of the PCR band corresponding to the bacterial 16S rRNA gene varied depending on the combination of primers and extraction kit. Primers  $\alpha$  produced the weakest PCR band in combination with kit A, but revealed a clearer band in combination with kits B and C. Primers  $\beta$  and  $\gamma$  were highly effective at producing conspicuous amplicons, regardless of the kit used (Fig. 2a).

In AB-treated butterflies, the bacterial signals (PCR bands) were very weak in intensity, irrespective of the type of primers and extraction kits. Hence, these butterflies harbour a very low bacterial biomass. Primers  $\alpha$  produced a second amplicon when used in combination with kits B and C. Sanger sequencing (Microsynth AG) confirmed that this amplicon originates from the insect 18S rRNA gene (Table S4). Interestingly, this band was not observed when using kit A (Fig. 2a).

For *P. brassicae* eggs (from untreated females), the PCR band corresponding to the bacterial 16S rRNA gene was extremely tenuous or absent in the analysed samples (Fig. 2b). Similar to the samples from AB-treated butterflies, primers  $\alpha$  produced an amplicon belonging to the insect 18S rRNA gene (Table S4). This second band was



**Figure 2** PCR amplification of the 16S rRNA gene from *Pieris brassicae* samples. PCR amplification using three different primers (Prim:  $\alpha$ ,  $\beta$  and  $\gamma$ ) targeting different regions of the 16S rRNA gene in DNA samples obtained with different extraction kits (A, B and C). Samples (1–10) consisted of (a) antibiotic-treated (red) and untreated (white) female butterflies and (b) eggs. Controls comprised negative and positive DNA extraction controls (–, +) and negative and positive PCR controls ((–), (+)). For some samples (\*) processed with primers  $\alpha$ , an amplicon of the insect 18S rRNA gene was detected. Expected size of the 16S amplicon is indicated by the positive controls.  $N = 10$  for each type of sample.

detected in samples processed with kits A and B, but not with kit C.

In summary, the intensity of the bacteria signal (PCR band) obtained from untreated butterflies was dependent on the combination of DNA extraction kit and primers used. In contrast, the bacteria signal intensity was very weak when analysing the butterfly eggs and AB-treated adults, regardless of the kit and primer pair used.

#### Effects of the DNA extraction kit on detection of bacteria taxa in *P. brassicae* butterflies

We sequenced bacterial PCR amplicons from untreated and AB-treated butterflies processed with the three different DNA extraction kits (Table 1) to assess whether the kit used affects which bacteria taxa are detected. Out of the two primers ( $\alpha$  and  $\gamma$ ) that can yield an amplicon of adequate size (<500 bp) for the Illumina MiSeq platform, we selected primers  $\gamma$  because they yielded a single amplicon of the 16S rRNA gene (compare Fig. 2).

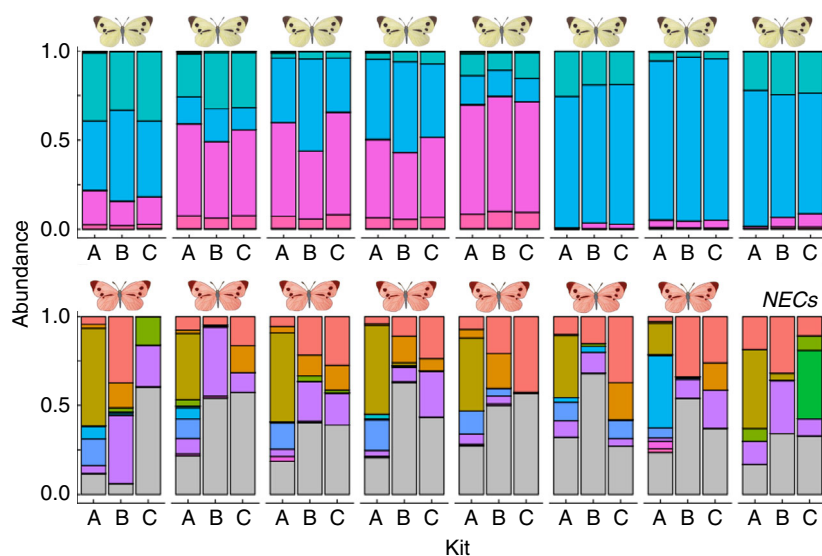
The sequencing of bacteria in *P. brassicae* adults revealed striking differences in the bacterial communities associated with untreated and AB-treated butterflies. In untreated *P. brassicae* adults, the community composition showed great homogeneity across samples processed with the different kits. Thus, the sequencing outcome of their bacterial community was not significantly affected by the kit used (Fig. 3). These samples were dominated by four

bacteria genera (*Gluconobacter*, *Lactococcus*, *Serratia* and *Yersinia*), which accounted for 98–99.9% of the reads. In contrast, the bacterial communities found in AB-treated butterflies and the NECs obtained with kits A, B and C were highly heterogeneous and varying with the kit used for processing the samples; they were dominated by bacterial taxa which are commonly reported as contaminants (Salter *et al.* 2014). The taxa *Pseudomonas* and *Acinetobacter* were present in all samples of AB-treated butterflies processed with the three kits.

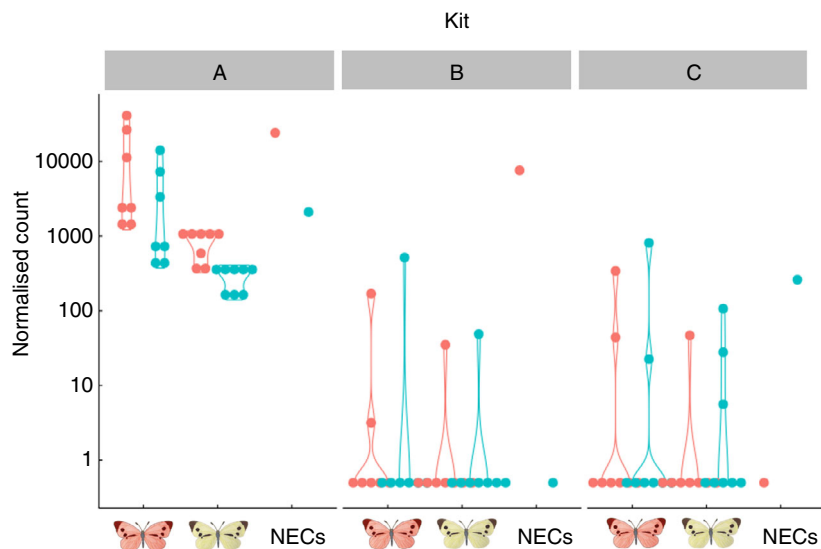
Furthermore, when analysing the differential abundance of bacteria genera in the differently extracted samples of control and AB-treated butterflies, a likelihood ratio test revealed increased counts of *Burkholderia* ( $P < 0.001$ ) and *Methylobacterium* ( $P = 0.023$ ) associated with kit A (Fig. 4). This result indicates that those members are contaminants belonging to kit A and are not part of the bacterial community of *P. brassicae* adults.

#### Effects of the 'kitome' on the characterization of bacterial communities in *P. brassicae* butterflies by dissimilarity and diversity indices

We further determined the dissimilarity of the bacterial communities in the differently processed butterfly samples and assessed the factors, which account for this. Since our analysis indicated the presence of bacterial contaminants in the kits, we included the detected bacteria



**Figure 3** Relative bacteria taxa abundance in untreated and AB-treated *Pieris brassicae* adults. The 12 most abundant bacteria genera are depicted: (■) *Acinetobacter*, (■) *Acquabacterium*, (■) *Burkholderia*, (■) *Chryseobacterium*, (■) *Enterococcus*, (■) *Ewingella*, (■) *Gluconobacter*, (■) *Lactococcus*, (■) *Methylobacterium*, (■) *Pseudomonas*, (■) *Serratia*, (■) *Yersinia*, (■) other (see Table S5 for proportions represented by these taxa). Each group of three stacked bars represents one individual butterfly processed with DNA extraction kit A, B or C. Negative extraction controls (NECs) correspond to mock samples processed in parallel using each kit.  $N = 8$  untreated adult butterflies (white) for control, 7 AB-treated adult butterflies (red) and 1 NEC per kit.



**Figure 4** Normalized counts of sequencing reads corresponding to ■ *Burkholderia* and ■ *Methylobacterium*. *Burkholderia* and *Methylobacterium* were the only taxa significantly associated (\*\*\*) with one of the DNA extraction kits (kit A) found in samples from control butterflies (white), antibiotic-treated butterflies (red) and negative extraction controls (NECs) *Burkholderia*:  $P < 0.001$ , *Methylobacterium*:  $P = 0.023$ , FDR-corrected. Counts were normalized by size factors accounting for variation in sequencing depth across samples. Size factors were estimated using the 'median of ratios' method described by equation 5 in Anders and Huber (2010). A modified geometric mean was used by taking the tenth root of the product of the non-zero counts (McMurdie and Holmes 2013).  $N = 7-8$  butterfly samples for each kit,  $N = 1$  NEC per kit, likelihood ratio test (LRT).

taxa ('kitomes') of NECs in this analysis. We conducted a PCoA ordination of Bray–Curtis dissimilarity (based on the bacteria genera detected and their abundance) followed by permutational analysis of multivariate homogeneity of group dispersions (Fig. 5).

Bacterial communities differed according to treatment regardless of the extraction kit ( $F = 14.3$ ,  $R^2 = 0.38$ ,  $P = 0.001$ ). The bacterial community of the pooled control butterfly samples differed significantly from both the one of the pooled AB-treated samples ( $F = 25.29$ ,  $R^2 = 0.37$ ,  $P = 0.003$ ) and from the pooled NECs ( $F = 11.70$ ,  $R^2 = 0.32$ ,  $P = 0.003$ ) (Fig. 5).

When further analysing the impact of treatment and kit on the dissimilarities, we found a significant effect of the AB treatment ( $F = 17.10$ ,  $R^2 = 0.38$ ,  $P = 0.001$ ) and kit ( $F = 3.25$ ,  $R^2 = 0.07$ ,  $P = 0.001$ ) and an effect of a kit by treatment interaction ( $F = 2.12$ ,  $R^2 = 0.10$ ,  $P = 0.005$ ) on the dissimilarities.

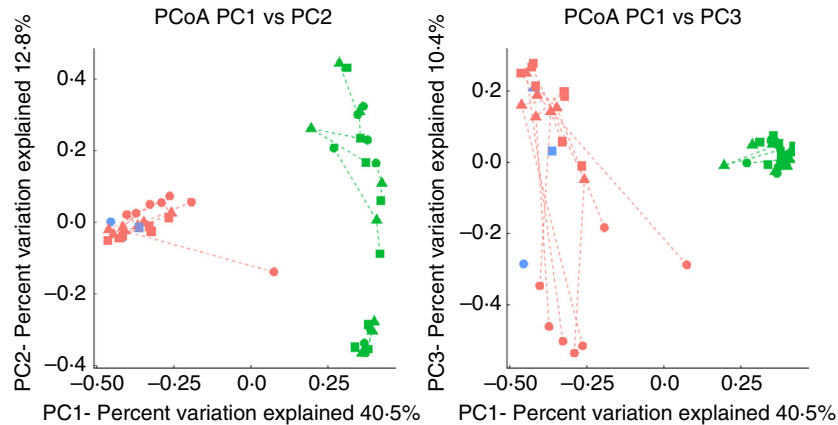
When considering only the samples from AB-treated butterflies, we found a significant effect of the type of kit ( $F = 4.78$ ,  $R^2 = 0.31$ ,  $P = 0.001$ ) on community dissimilarities. Hence, samples of AB-treated butterflies processed with kit A differed from kit B ( $F = 6.24$ ,  $R^2 = 0.34$ ,  $P = 0.003$ ) and kit C ( $F = 5.96$ ,  $R^2 = 0.31$ ,  $P = 0.003$ ). In contrast, when considering only the samples from untreated butterflies, there was no effect of the kit ( $F = 2.66$ ,  $R^2 = 0.08$ ,  $P = 0.063$ ).

To characterize the bacterial communities by their alpha diversities, we compared the Shannon indices of

the bacterial communities detected in the differently processed samples of control and AB-treated butterflies as well as in the NECs. We used bacteria genera and their abundance for calculating the index. The alpha diversities varied according to treatment (Fig. 6). The bacterial communities detected in AB-treated butterflies and NECs showed a greater alpha diversity than those in untreated butterflies. Thus, samples from untreated butterflies significantly differed in their diversity from both the AB-treated butterflies ( $T = -10.5$ ,  $df = 46$ ,  $P < 0.01$ ) and NECs ( $T = -4.84$ ,  $df = 46$ ,  $P < 0.01$ ). There was no significant effect of kit nor of a kit by treatment interaction on the alpha diversity index (Fig. S1).

In summary, the sequencing of amplicons obtained with primers  $\gamma$  revealed that the detected bacterial community composition of samples with high bacterial abundance remained consistent across kits and was dominated only by a few taxa. In the presence of such dominant members, kit-associated bacteria occurring in very low abundances could not be detected upon sequencing. Consequently, the alpha diversity of the bacterial community of untreated control butterflies was very low. In contrast, the sequencing outcome of the community composition in samples with low bacterial abundance (AB-treated adults) was dependent on the extraction kit and the bacterial contaminants therein. Many bacteria genera of low dominance were detectable in these samples upon sequencing, resulting in a high alpha diversity of the respective communities. Hence, the 'kitome' shaped the



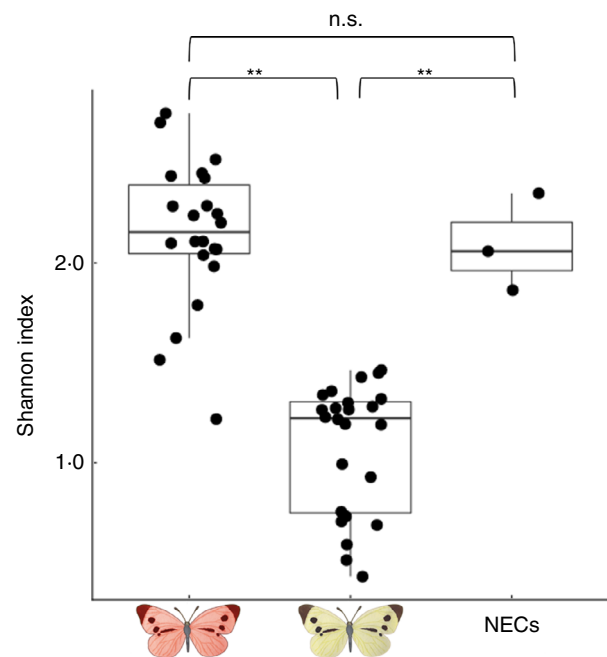


**Figure 5** Principal coordinate analysis (PCoA) ordination of Bray–Curtis dissimilarity of butterfly-associated bacterial communities in *Pieris brassicae* adults. AB-treated butterflies (red), untreated butterflies (green) and negative extraction controls (NECs) (blue) processed with different DNA extraction kits: A (circles), B (triangles) and C (squares). The dashed lines connect samples from the same individual extracted with different kits. There was a significant effect of the antibiotic (AB) treatment ( $F = 17.10$ ,  $R^2 = 0.38$ ,  $P = 0.001$ ) and kit ( $F = 3.25$ ,  $R^2 = 0.07$ ,  $P = 0.001$ ), and AB\*kit interaction ( $F = 2.12$ ,  $R^2 = 0.10$ ,  $P = 0.005$ ).  $N = 7$ – $8$  AB-treated and control butterflies, 1 negative extraction control (NEC) per kit, PERMANOVA.

alpha diversity index of samples with a low insect-associated bacterial abundance.

## Discussion

Assessing the most appropriate DNA extraction technique is fundamental when investigating the bacterial community in an unexplored ecosystem. Here, we analysed how the selection of DNA extraction kit and bacteria therein affect the detection of bacteria in lepidopteran samples with different bacterial abundance: eggs, AB-treated and untreated butterflies. We took *P. brassicae* as a model and found, as expected, only a very low abundance of bacteria associated with *P. brassicae* eggs and AB-treated butterflies, but consistent detection of bacteria in conspicuous abundance in untreated *P. brassicae* adults. In this latter type of sample, the bacteria signal intensity obtained by the PCR analysis varied with the type of DNA extraction kit and the primers used, suggesting differences in the extraction efficiency of the kits and/or their compatibility with the primers. However, Illumina (MiSeq) sequencing revealed that the type of kit used for extraction of bacterial DNA hardly exerts any effects on the taxon identification of the detected bacterial community in those lepidopteran samples with a high bacterial abundance (untreated butterflies). In striking contrast, the kit determined the bacterial community composition in the samples with a low bacterial abundance; the ‘kitome’ shaped the bacterial community composition and significantly affected the determined alpha diversity index of the bacterial community.



**Figure 6** Alpha diversity (Shannon index) of butterfly-associated bacterial communities. Samples processed with the different kits were pooled for the analysis based on the insect treatment: antibiotic-treated (red) and untreated (white). The negative extraction controls (NECs) processed in parallel with the true samples during DNA extraction were pooled. For each boxplot: inner line = median, box = 25–75% quartiles, whiskers = highest and lowest values no greater or lower than the first or third quartile plus 1.5 times the interquartile range.  $N = 24$  control, 21 AB, 3 NECs. Asterisks indicate significant differences ( $P < 0.01$ ), n.s. indicates no significant difference, Wald tests, based on bacteria genera.

How to explain that the PCR analysis revealed clear effects of the kit on the bacteria signal intensity in combination with the type of primers used? In addition to primer pair  $\gamma$ , primer pair  $\beta$  also functioned kit independently and provided a strong bacteria signal in samples from control adults. Primers  $\alpha$ , however, gave a clear signal when samples were processed with kits B and C, and only a weak signal when samples were processed with kit A. The reason behind the weak signal given by primers  $\alpha$  in samples processed with kit A might be the presence of DNA extraction-related PCR inhibitors differentially affecting primer annealing (Schrader *et al.* 2012). This idea is supported by the weaker signal given by the positive extraction control in comparison to the positive PCR control (compare '+' vs '(+)’ kit A, Fig. 2a), showing that the extraction with kit A reduced the efficacy of the PCR with primers  $\alpha$ , but not with primers  $\beta$  and  $\gamma$ . This kit-mediated effect might be caused by differences in the quantitative and/or qualitative chemical composition of the kits, affecting the purity of the DNA product. However, the exact chemical composition of the kit reagents is not publicly available. Whether such PCR inhibition would be observed when using a more sensitive DNA polymerase remains to be tested. We chose the JumpStart™ DNA polymerase due to its high performance at a price, which is representative of most DNA polymerases used for the amplification of the 16S gene. Another potential explanation for the weaker signal shown by kit A in combination with primers  $\alpha$  is a low bacterial DNA yield in relation to the total DNA yield (from insects and their bacteria) obtained with this kit. We measured the total DNA yield obtained by the tested kits and found that kit A produced a significantly lower yield per sample than kits B and C (Fig. S2), thus showing a reduced extraction efficiency. DNA yield is used as a quality parameter for DNA extraction kits (Pollock *et al.* 2018). Different DNA yields obtained by different kits may be due to their varying capacity to disrupt cells (de Bruin and Birnboim 2016), especially if the sample contains Gram-positive bacteria and endospores, which are less prone to lysis (Pollock *et al.* 2018). Thus, poor DNA extraction can produce an incomplete coverage of the true bacterial community during downstream PCR amplification (Ariefdjohan *et al.* 2010). Overall, our results indicate that the type of kit can affect the efficacy of the 16S primers.

Interestingly, we found that primers  $\alpha$  also amplify 18S rRNA (insect) amplicons. This renders them less suitable for (MiSeq) sequencing of bacterial communities in *P. brassicae* since it demands an extra amplicon purification process prior to (MiSeq) sequencing. Hence, primers  $\gamma$  were chosen for the community sequencing due to the single amplicon they yield and the explicit

recommendation by Illumina to use these primers for the MiSeq platform (Illumina 2013; Klindworth *et al.* 2013). Remarkably, the use of these primers produced very low numbers of plant and insect mitochondrial 16S reads (Table S3). This result may vary when using these primers on other lepidopteran species and life stages (e.g. folivorous larvae with high plant content in their gut). In such cases, use of insect- or plant-specific blocking primers might greatly favour the amplification of bacterial sequences over insect and plant sequences (Vestheim and Jarman 2008).

The simplicity of the bacterial community in untreated *P. brassicae* butterflies is likely explaining the lack of significant differences in the composition of the detected adult-associated bacterial community when using different DNA extraction kits. The majority of studies showing a large effect of DNA extraction methods on the retrieved bacterial community composition are focused on animal samples containing robust, taxon-rich bacterial communities (Scupham *et al.* 2007; Yuan *et al.* 2012; Henderson *et al.* 2013; Larsen *et al.* 2015; Burbach *et al.* 2016; Weber *et al.* 2017; Ketchum *et al.* 2018). Lepidoptera, in contrast, are known to be colonized by rather few bacteria taxa (Paniagua Voirol *et al.* 2018). Our results indicate that the type of kit did not influence the detection of bacteria taxa when handling samples with conspicuous bacteria abundance, but low community diversity. Samples from AB-treated butterflies and NECs had higher bacterial diversity (Shannon indices) than the control adult samples. This calculated higher diversity turned out to be due to several kit contaminating bacteria, which were detected in these types of samples because of the highly reduced abundance of bacteria associated with the butterflies. The presence of several bacterial contaminants in low and similar abundances as well as the lack of highly dominant taxa contributed to the high Shannon index, which is the higher the more taxa are present in similar abundances. We suggest that the bacterial kit contaminants present in samples from untreated butterflies were marginally or not amplified by PCR, and thus neither identified by the sequencing analysis, presumably due to the high abundance of the bacteria associated with these Lepidoptera samples.

Understanding the impact of bacteria on the biology of organisms associated with them is not only a challenge for entomologists, but for all biologists and physicians, who need to analyse samples with low bacterial biomass such as mammalian placenta or blood plasma. This is challenging because commonly used methods for samples with high microbial abundance (e.g. human gut and faeces) do not generate reliable readouts for samples with low microbial abundance (Weiss *et al.* 2014). There is an increasing evidence that analyses of samples with low

microbial biomass are highly susceptible to produce biased results, overestimating the impact of bacterial taxa, which are members of the bacterial community present in the kit, that is, of the 'kitome' or other contaminants (Weiss *et al.* 2014). Yet, many studies of bacterial communities lack sequencing of negative controls or descriptions of contamination removal methods (Salter *et al.* 2014; Glassing *et al.* 2016). Results like ours, showing higher bacteria diversity in the samples from AB-treated butterflies in comparison to control butterflies, could be misleading if no negative controls would have been included in the analysis.

The analysis of negative controls should be a basic requirement in the study of bacterial associates of Lepidoptera, especially when sampling life stages containing low bacteria biomass such as lepidopteran eggs or larvae (Hammer *et al.* 2017; Paniagua Voirol *et al.* 2020), or when analysing the effects of a treatment with ABs on insect-associated communities. Interestingly, many of the core bacteria genera reported in Lepidoptera, such as *Pseudomonas*, *Bacillus*, *Enterococcus* and *Acinetobacter* (Paniagua Voirol *et al.* 2018), have also been reported as common laboratory contaminant genera (Salter *et al.* 2014; Glassing *et al.* 2016). Even though this does not imply these bacteria are mere artefacts, appropriate controls are needed to confirm their association with insects. Since *Pseudomonas* and *Acinetobacter* were present in negative controls processed by all three kits tested here, we cannot fully exclude that these bacteria originate from other sources of contamination than the kit, such as plastic consumables (Motley *et al.* 2014), ultrapure water (McFeters *et al.* 1993; McAlister *et al.* 2002), PCR reagents (Grahn *et al.* 2003) or the laboratory environment (Witt *et al.* 2009).

Although *Burkholderia* and *Methylobacterium* were present in all samples processed by the different kits, their counts were significantly higher in samples processed with kit A (Fig. 4). Hence, there was a correlation between usage of kit A and these two bacteria genera. This, along with other taxa found in NECs, strongly indicates that DNA extraction kits contain bacterial DNA that influences the results of community sequencing if samples contain low bacterial presence.

In summary, the DNA extraction kit can affect the characterization of bacterial associates of Lepidoptera, especially when samples contain low bacterial abundance. In contrast to untreated butterflies, the detected bacterial community of AB-treated butterflies with reduced bacterial abundance was largely shaped by kit-associated contamination. Characterizing bacterial communities by alpha diversity indices needs in parallel the sequencing of NECs to prevent that the 'kitome' shapes statements on the diversity of the actual bacterial community under study.

## Acknowledgements

We thank Arne Weinhold (Freie Universität Berlin) for his contribution in the sequencing of insect-derived PCR amplicons. We thank Ute Braun (Freie Universität Berlin) for her assistance in growing plants and rearing the insects. This study was financed by the German Research Foundation (Deutsche Forschungsgemeinschaft) (Collaborative Research Centre 973, project B4, [www.sfb973.de](http://www.sfb973.de)). Open access funding enabled and organized by Projekt DEAL.

## Conflict of Interest

No conflict of interest is declared.

## Authors' contribution

L.R.P.V. and M.H. conceptualized the study. L.R.P.V., G.V. and M.Y. performed the experiments. P.R.J. performed the sequencing analysis. L.R.P.V. and M.H. wrote the first draft of the manuscript. All authors contributed to later versions of the manuscript and agreed with the final version.

## Data Availability Statement

Illumina sequencing data were deposited in SRA database under the bioproject accession PRJNA640344 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA640344>).

## References

- Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.
- Andreolli, M., Lampis, S., Poli, M., Gullner, G., Biró, B. and Vallini, G. (2013) Endophytic *Burkholderia fungorum* DBT1 can improve phytoremediation efficiency of polycyclic aromatic hydrocarbons. *Chemosphere* **92**, 688–694.
- Ariefdjohan, M.W., Savaiano, D.A. and Nakatsu, C.H. (2010) Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutr J* **9**, 23.
- Baker, G.C., Smith, J.J. and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**, 541–555.
- Ben-Dov, E., Shapiro, O.H., Siboni, N. and Kushmaro, A. (2006) Advantage of using inosine at the 3' termini of 16S rRNA gene universal primers for the study of microbial diversity. *Appl Environ Microbiol* **72**, 6902–6906.
- Bringel, F. and Couee, I. (2015) Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. *Front Microbiol* **6**, 486.

- Bukin, Y.S., Galachyants, Y.P., Morozov, I.V., Bukin, S.V., Zakharenko, A.S. and Zemskaya, T.I. (2019) The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* **6**, 190007.
- Burbach, K., Seifert, J., Pieper, D.H. and Camarinha-Silva, A. (2016) Evaluation of DNA extraction kits and phylogenetic diversity of the porcine gastrointestinal tract based on Illumina sequencing of two hypervariable regions. *MicrobiologyOpen* **5**, 70–82.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J. and Holmes, S.P. (2016a) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583.
- Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. and Holmes, S.P. (2016b) Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F100Research* **5**, 1492.
- Castelino, M., Eyre, S., Moat, J., Fox, G., Martin, P., Ho, P., Upton, M. and Barton, A. (2017) Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Microbiol* **17**, 23.
- Chen, B., Teh, B.S., Sun, C., Hu, S., Lu, X., Boland, W. and Shao, Y. (2016) Biodiversity and activity of the gut microbiota across the life history of the insect herbivore *Spodoptera littoralis*. *Sci Rep* **6**, 29505.
- Choo, J.M., Leong, L.E. and Rogers, G.B. (2015) Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep* **5**, 16350.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A. et al. (2014) Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633–642.
- Cuív, P.Ó., De Cárcer, D.A., Jones, M., Klaassens, E.S., Worthley, D.L., Whitehall, V.L., Kang, S., McSweeney, C.S. et al. (2011) The effects from DNA extraction methods on the evaluation of microbial diversity associated with human colonic tissue. *Microb Ecol* **61**, 353–362.
- Dalla-Costa, L.M., Morello, L.G., Conte, D., Pereira, L.A., Palmeiro, J.K., Ambrosio, A., Cardozo, D., Krieger, M.A. et al. (2017) Comparison of DNA extraction methods used to detect bacterial and yeast DNA from spiked whole blood by real-time PCR. *J Microbiol Methods* **140**, 61–66.
- de Bruin, O.M. and Birnboim, H. (2016) A method for assessing efficiency of bacterial cell disruption and DNA release. *BMC Microbiol* **16**, 197.
- Dohet, L., Gregoire, J.-C., Berasategui, A., Kaltenpoth, M. and Biedermann, P.H. (2016) Bacterial and fungal symbionts of parasitic *Dendroctonus* bark beetles. *FEMS Microbiol Ecol* **92**(9), <https://doi.org/10.1093/femsec/fiw129>.
- Douglas, A.E. (2015) Multiorganismal insects: diversity and function of resident microorganisms. *Annu Rev Entomol* **60**, 17–34.
- Eldridge, D.J., Delgado-Baquerizo, M., Travers, S.K., Val, J., Oliver, I., Hamonts, K. and Singh, B.K. (2017) Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. *Ecology* **98**, 1922–1931.
- Engl, T., Bodenstern, B. and Strohm, E. (2016) Mycobiota in the brood cells of the European beewolf, *Philanthus triangulum* (Hymenoptera: Crabronidae). *Eur J Entomol* **113**, 271–277.
- Evans, J.S., Lopez-Legentil, S. and Erwin, P.M. (2018) Comparing two common DNA extraction kits for the characterization of symbiotic microbial communities from ascidian tissue. *Microbes Environ* **33**, 435–439.
- Fadrosch, D.W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R.M. and Ravel, J. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6.
- Feldhaar, H. (2011) Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol* **36**, 533–543.
- Ferrari, J. and Vavre, F. (2011) Bacterial symbionts in insects or the story of communities affecting communities. *Philos Trans R Soc Lond B Biol Sci* **366**, 1389–1400.
- Flórez, L.V., Scherlach, K., Gaube, P., Ross, C., Sitte, E., Hermes, C., Rodrigues, A., Hertweck, C. et al. (2017) Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. *Nature Commun* **8**, 15172.
- Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J. and Cotter, P.D. (2016) 16S rRNA gene sequencing of mock microbial populations—impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* **16**, 123.
- Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C. et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**, 298–308.
- Glassing, A., Dowd, S.E., Galandiuk, S., Davis, B. and Chiodini, R.J. (2016) Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog* **8**, 24.
- Grahn, N., Olofsson, M., Ellnebo-Svedlund, K., Monstein, H.J. and Jonasson, J. (2003) Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol Lett* **219**, 87–91.
- Hammer, T.J., Janzen, D.H., Hallwachs, W., Jaffe, S.P. and Fierer, N. (2017) Caterpillars lack a resident gut microbiome. *Proc Natl Acad Sci U S A* **114**, 9641–9646.
- Hammer, T.J., McMillan, W.O. and Fierer, N. (2014) Metamorphosis of a butterfly-associated bacterial community. *PLoS One* **9**, e86995.
- Hansen, A.K. and Moran, N.A. (2014) The impact of microbial symbionts on host plant utilization by herbivorous insects. *Mol Ecol* **23**, 1473–1496.

- Henderson, G., Cox, F., Kittelmann, S., Miri, V.H., Zethof, M., Noel, S.J., Waghorn, G.C. and Janssen, P.H. (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS One* **8**, e74787.
- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M. et al. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207.
- Illumina, I. (2013) 16S Metagenomic sequencing library preparation, pp. 1–28. Illumina.
- Kaltenpoth, M., Yildirim, E., Gürbüz, M.F., Herzner, G. and Strohm, E. (2011) Refining the roots of the beewolf-*Streptomyces* symbiosis: antennal symbionts in the rare genus *Philanthinus* (Hymenoptera, Crabronidae). *Appl Environ Microbiol* **78**, 822–827.
- Ketchum, R.N., Smith, E.G., Vaughan, G.O., Phippen, B.L., McParland, D., Al Mansoori, N., Carrier, T.J., Burt, J.A. et al. (2018) DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Front Mar Sci* **5**, 255.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acid Res* **41**, e1.
- Koga, R. and Moran, N.A. (2014) Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J* **8**, 1237–1246.
- Kuhn, M., Weston, S., Wing, J., Forester, J. and Thaler, T. (2013) Contrast: a collection of contrast methods. R package version 0.19.
- Larsen, A.M., Mohammed, H.H. and Arias, C.R. (2015) Comparison of DNA extraction protocols for the analysis of gut microbiota in fishes. *FEMS Microbiol Lett* **362**, fnu031.
- Leonhardt, S.D. and Kaltenpoth, M. (2014) Microbial communities of three sympatric Australian stingless bee species. *PLoS One* **9**, e105718.
- Lievens, B., Hallsworth, J.E., Pozo, M.I., Belgacem, Z.B., Stevenson, A., Willems, K.A. and Jacquemyn, H. (2015) Microbiology of sugar-rich environments: diversity, ecology and system constraints. *Environ Microbiol* **17**, 278–298.
- Lindh, J.M. and Lehane, M.J. (2011) The tsetse fly *Glossina fuscipes fuscipes* (Diptera: *Glossina*) harbours a surprising diversity of bacteria other than symbionts. *Antonie Van Leeuwenhoek* **99**, 711–720.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550.
- Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Avila, J.A., López-Guerrero, J.A. and Aguilar-Quesada, R. (2016) DNA source selection for downstream applications based on DNA quality indicators analysis. *Biopreserv Biobank* **14**, 264–270.
- Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J.C., Soulas, G. and Catroux, G. (2001) DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* **67**, 2354–2359.
- McAlister, M.B., Kulakov, L.A., O'Hanlon, J.F., Larkin, M.J. and Ogden, K.L. (2002) Survival and nutritional requirements of three bacteria isolated from ultrapure water. *J Ind Microbiol Biotechnol* **29**, 75–82.
- McMurdie, P.J. and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217.
- McFeters, G.A., Broadaway, S.C., Pyle, B.H. and Egozy, Y. (1993) Distribution of bacteria within operating laboratory water purification systems. *Appl Environ Microbiol* **59**, 1410–1415.
- Méthé, B.A., Hiorns, W.D. and Zehr, J.P. (1998) Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George and six other Adirondack lakes. *Limnol Oceanogr* **43**, 368–374.
- Motley, S.T., Picuri, J.M., Crowder, C.D., Minich, J.J., Hofstadler, S.A. and Eshoo, M.W. (2014) Improved multiple displacement amplification (iMDA) and ultraclean reagents. *BMC Genom* **15**, 443.
- Oksanen, J., Blanchet, F., Guillaume, F., Kindt, R., Legendre, P., McGinn, D. and Wagner, H. (2018) Vegan: community ecology package. R package version 2.5. 2–5. <https://CRAN.R-project.org/package=vegan>
- Paniagua Voirol, L.R., Frago, E., Kaltenpoth, M., Hilker, M. and Fatouros, N.E. (2018) Bacterial symbionts in Lepidoptera: their diversity, transmission, and impact on the host. *Front Microbiol* **9**, 556.
- Paniagua Voirol, L.R., Valsamakis, G., Lortzing, V., Weinhold, A., Johnston, P.R., Fatouros, N., Kunze, R. and Hilker, M. (2020) Plant responses to insect eggs are not induced by egg-associated microbes, but by a secretion attached to the eggs. *Plant Cell Environ* **43**, 1815–1826.
- Pankewitz, F., Zollmer, A., Hilker, M. and Graser, Y. (2007) Presence of *Wolbachia* in insect eggs containing antimicrobially active anthraquinones. *Microb Ecol* **54**, 713–721.
- Pendleton, K.M., Erb-Downward, J.R., Bao, Y., Branton, W.R., Falkowski, N.R., Newton, D.W., Huffnagle, G.B. and Dickson, R.P. (2017) Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. *Am J Respir Crit* **196**, 1610–1612.
- Phalnikar, K., Kunte, K. and Agashe, D. (2018) Dietary and developmental shifts in butterfly-associated bacterial communities. *R Soc Open Sci* **5**, 171559.
- Pollock, J., Glendinning, L., Wisedchanwet, T. and Watson, M. (2018) The madness of microbiome: attempting to find consensus 'best practice' for 16S microbiome studies. *Appl Environ Microbiol* **84**, e02627–e12617.

- R Core Team (2018) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ravenscraft, A., Berry, M., Hammer, T., Peay, K. and Boggs, C. (2019) Structure and function of the bacterial and fungal gut microbiota of neotropical butterflies. *Ecol Monograph* **89**, e01346.
- Riesenfeld, C.S., Schloss, P.D. and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**, 525–552.
- Salem, H., Kreutzer, E., Sudakaran, S. and Kaltenpoth, M. (2013) Actinobacteria as essential symbionts in firebugs and cotton stainers (Hemiptera, Pyrrhocoridae). *Environ Microbiol* **15**, 1956–1968.
- Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J. et al. (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* **12**, 87.
- Schrader, C., Schielke, A., Ellerbroek, L. and Johne, R. (2012) PCR inhibitors—occurrence, properties and removal. *J Appl Microbiol* **113**, 1014–1026.
- Scupham, A.J., Jones, J.A. and Wesley, I.V. (2007) Comparison of DNA extraction methods for analysis of turkey cecal microbiota. *J Appl Microbiol* **102**, 401–409.
- Staudacher, H., Kaltenpoth, M., Breeuwer, J.A., Menken, S.B., Heckel, D.G. and Groot, A.T. (2016) Variability of bacterial communities in the moth *Heliothis virescens* indicates transient association with the host. *PLoS One* **11**, e0154514.
- Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A. et al. (2017) A communal catalogue reveals earth's multiscale microbial diversity. *Nature* **551**, 457.
- Torsvik, V. and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5**, 240–245.
- Unno, T. (2015) Bioinformatic suggestions on MiSeq-based microbial community analysis. *J Microbiol Biotechnol* **25**, 765–770.
- Vestheim, H. and Jarman, S.N. (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Front Zool* **5**, 12.
- Weber, L., De Force, E. and Apprill, A. (2017) Optimization of DNA extraction for advancing coral microbiota investigations. *Microbiome* **5**, 18.
- Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**, 697–703.
- Weiss, S., Amir, A., Hyde, E.R., Metcalf, J.L., Song, S.J. and Knight, R. (2014) Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol* **15**, 564.
- Willis, A.D. (2019) Rarefaction, alpha diversity, and statistics. *Front Microbiol* **10**, 2407.
- Wintzingerode, F., Göbel, U.B. and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**, 213–229.
- Witt, N., Rodger, G., Vandesompele, J., Benes, V., Zumla, A., Rook, G.A. and Huggett, J.F. (2009) An assessment of air as a source of DNA contamination encountered when performing PCR. *J Biomol Technol* **20**, 236.
- Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z. and Forney, L.J. (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, e33865.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Materials and methods.

**Data S2.** Results.

**Table S1.** Amplicon sequence variants (ASV) per sample before agglomeration at the genus level.

**Table S2.** Sequencing library size per sample.

**Table S3.** Percentage of reads from chloroplast and insect/plant mitochondria per sample.

**Table S4.** Sequence of the insect 18S rRNA gene amplified by 515F/806R primers.

**Table S5.** Proportion of the *Pieris brassicae* bacterial community represented by the 12 most abundant taxa.

**Figure S1.** Alpha diversity (Shannon index) of butterfly-associated bacterial communities extracted with different DNA extraction kits.

**Figure S2.** Total DNA yield per sample obtained by using three different extraction kits.