



**HAL**  
open science

## **Nested PCR to optimize rpoB metabarcoding for lowconcentration and host-associated bacterial DNA**

Laëtitia Leclerc, Gabryelle Agoutin, Thierry Brévault, Antony Champion, Jean Mainguy, Nicolas Nègre, Sudeeptha Yainna, Géraldine Pascal, Sophie Gaudriault, Jean-Claude Ogier

### ► **To cite this version:**

Laëtitia Leclerc, Gabryelle Agoutin, Thierry Brévault, Antony Champion, Jean Mainguy, et al.. Nested PCR to optimize rpoB metabarcoding for lowconcentration and host-associated bacterial DNA. *Microbiology Spectrum*, 2025, 13 (9), pp.e01417-25. <10.1128/spectrum.01417-25>. <hal-05220858>

**HAL Id: hal-05220858**

**<https://hal.inrae.fr/hal-05220858v1>**

Submitted on 24 Aug 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Nested PCR to optimize *rpoB* metabarcoding for low-concentration and host-associated bacterial DNA

Laëtitia Leclerc,<sup>1</sup> Gabryelle Agoutin,<sup>2</sup> Thierry Brévault,<sup>3</sup> Antony Champion,<sup>4</sup> Jean Mainguy,<sup>5,6</sup> Nicolas Nègre,<sup>1</sup> Sudeeptha Yainna,<sup>1</sup> Géraldine Pascal,<sup>2</sup> Sophie Gaudriault,<sup>1</sup> Jean-Claude Ogier<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 12.

**ABSTRACT** Housekeeping genes have proven to be effective taxonomic markers for characterizing bacterial microbiota in short-read amplicon metabarcoding studies. A region of the *rpoB* gene has been shown to minimize Operational Taxonomic Unit (OTU) overestimation bias with a high degree of accuracy, providing better species-level taxonomic resolution (J.-C. Ogier, S. Pagès, M. Galan, M. Barret, and S. Gaudriault, S., BMC Microbiol. 19:171, <https://doi.org/10.1186/s12866-019-1546-z>). However, the primers for *rpoB* are highly degenerate, leading to potential problems in the amplification of bacterial DNA present at low concentration in the sample or embedded within eukaryotic matrices, as for the host-associated microbiota. We addressed these limitations by using a two-step PCR approach to optimize the *rpoB* procedure. The first PCR amplifies a region of the *rpoB* gene with the outer primers, and the second PCR then uses primers incorporating Illumina adapters, the inner primers, to amplify the taxonomic marker for metabarcoding. We first used *in silico* approaches to evaluate the universality of the outer and inner *rpoB* primers. We then tested the nested *rpoB* PCR method on commercial mock samples. The nested PCR approach increased amplification efficiency for dilute samples without biasing the bacterial composition of the mock sample revealed by metabarcoding relative to single-step PCR. We also tested the nested *rpoB* PCR method on field-collected samples of insects. The nested PCR outperformed single-step PCR, increasing amplification efficiency for bacterial DNA present at low concentrations (insect oral secretions) or embedded in eukaryotic DNA matrices (insect larvae). This method provides a promising new strategy for characterizing host-associated microbiota.

**IMPORTANCE** In this study, we improved an existing method for characterizing host-associated bacterial microbiota using a genetic marker targeting a specific subregion of the *rpoB* gene. The two-step PCR approach enables more efficient amplification of bacterial DNA at low concentrations or within complex host DNA matrices, such as those found in insects. We optimized the cycle number in each PCR step to minimize bias in bacterial community composition, validated through tests on mock communities. This method provides a robust and reliable solution for analyzing host-associated microbiota, enhancing amplification efficiency and improving result accuracy.

**KEYWORDS** metabarcoding, host-associated microbiota, *rpoB*, nested PCR, insect, *Spodoptera frugiperda*

Microbial communities are frequently characterized by polymerase chain reaction (PCR) amplification followed by next-generation sequencing (NGS), particularly through short-read metabarcoding approaches, typically targeting variable regions of the 16S rRNA gene, such as the V3-V4 regions. However, short-amplicon sequencing

**Editor** John M. Chaston, Brigham Young University, Provo, Utah, USA

Address correspondence to Jean-Claude Ogier, [jean-claude.ogier@umontpellier.fr](mailto:jean-claude.ogier@umontpellier.fr).

The authors declare no conflict of interest.

See the funding table on p. 13.

**Received** 8 May 2025

**Accepted** 22 July 2025

**Published** 12 August 2025

Copyright © 2025 Leclerc et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

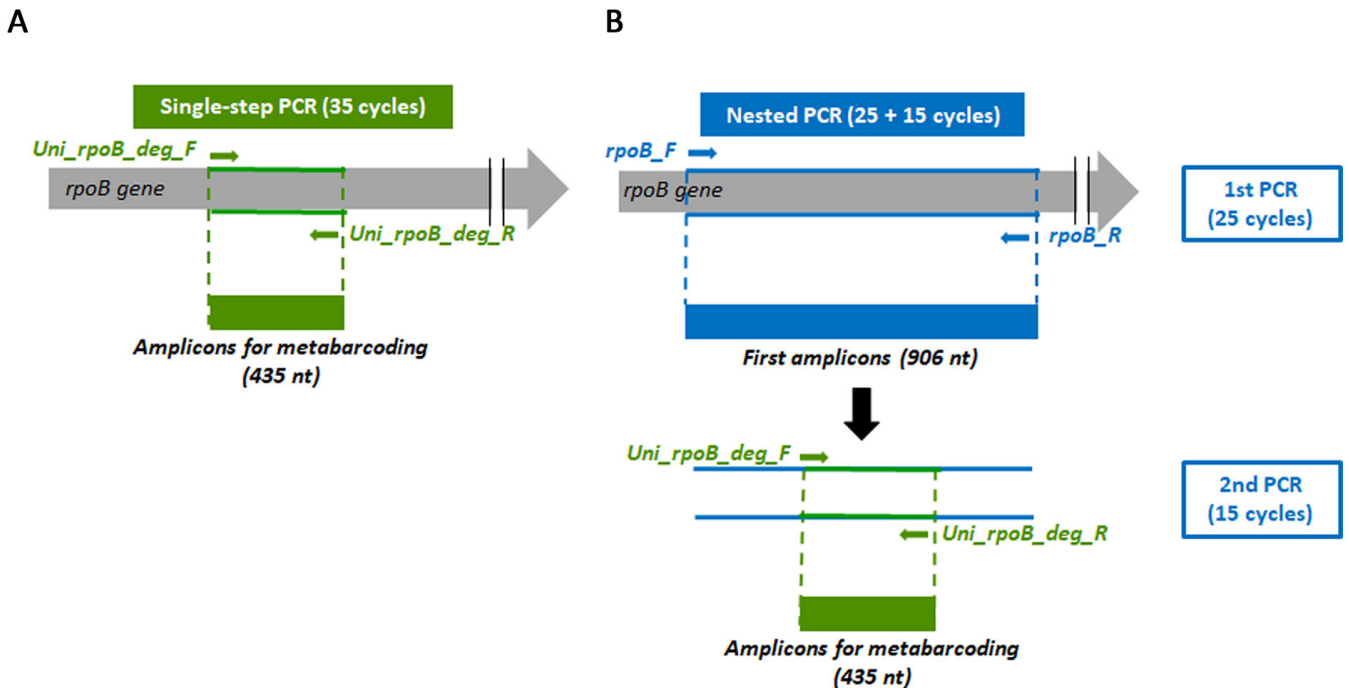
studies based on the 16S rRNA gene are subject to certain limitations, particularly in terms of taxonomic resolution, which is often restricted to the genus and family levels (1, 2). Full-length 16S rRNA gene sequencing with PacBio technology has been shown to improve taxonomic resolution in human microbiome samples (3), but its use is currently limited due to the higher cost required to obtain an equivalent number of reads per sample.

Housekeeping genes encoding conserved proteins ubiquitous in the bacterial kingdom provide a possible alternative to the short amplicon of the 16S rRNA gene. These genes evolve much more rapidly than the 16S rRNA gene and can therefore be used to differentiate between bacterial lineages (4). Moreover, unlike the 16S rRNA gene, housekeeping genes are generally present as a single copy in genomes, minimizing biases in diversity assessment and providing a more accurate representation of bacterial community composition. Housekeeping genes have proven to be effective for the characterization of bacterial diversity, particularly at the species level, across various ecosystems (4–8). The *rpoB* marker, in particular, has been successfully used in metabarcoding analyses of samples rich in Pseudomonadota (formerly Proteobacteria), such as entomopathogenic nematodes, and has been shown to have a finer taxonomic resolution than the 16S (V3-V4) marker (5, 9). In particular, phylogenies based on the same *rpoB* subregion are highly concordant with core genome phylogenies (10), indicating that this subregion is a robust marker for precise bacterial identification.

One key challenge in the metabarcoding of microbial communities is detecting prokaryotic DNA in samples with very low bacterial DNA concentrations, or in samples in which eukaryotic DNA predominates, as in the host-associated microbiota. In most cases, physically separating microbes from host tissues is difficult or even impossible, making global DNA extraction the only option. However, global DNA extraction typically results in a low microbial-to-host DNA ratio, limiting the efficiency of amplification with bacterial primers. This issue is further exacerbated by the use of primers targeting housekeeping genes, such as *rpoB*, which has a highly variable sequence, necessitating the use of degenerate bases to maximize species coverage. This may increase cross-reactivity with eukaryotic DNA, further hampering target amplification. Finally, in metabarcoding protocols, the long sequence and GC-rich regions of barcoding primers (about 60 nucleotides, including the target region, index sequences, and Illumina adapters) reduce the specificity of hybridization, making it even more challenging to amplify bacterial DNA from a host DNA matrix.

Nested PCR can be used to increase amplification sensitivity. In nested PCR approaches, there are two successive reactions, the first to amplify a larger region, and the second to amplify a smaller region nested within the first amplicon. Nested PCR is particularly useful for samples with low target DNA concentrations (11), but this method has only rarely been used in microbial ecology (12, 13), and its potential for metabarcoding has yet to be explored.

In this study, we evaluated the potential of a nested PCR approach combined with a metabarcoding analysis with *rpoB* primers for characterizing bacterial communities in samples with low bacterial DNA concentrations or in which the bacterial DNA is embedded in a eukaryotic matrix. For validation of the efficacy of this strategy, we compared the performance of the single-step and nested *rpoB* PCR methods on a set of mock samples. We then compared the efficiency of the nested *rpoB* PCR approach to that of single-step PCR on biological samples consisting of oral secretions (OS) from the lepidopteran *Spodoptera frugiperda*, which contain low concentrations of bacterial DNA, and *S. frugiperda* larvae in which the bacterial DNA is diluted in large amounts of eukaryotic DNA.



**FIG 1** Protocol for the amplification of a region of the *rpoB* gene by (A) single-step PCR and (B) nested PCR strategies. The positions of the primers are indicated by green arrows (*Uni\_rpoB\_deg\_F/R* primers) and blue arrows (*rpoB\_F/R* primers). *nt*, nucleotides.

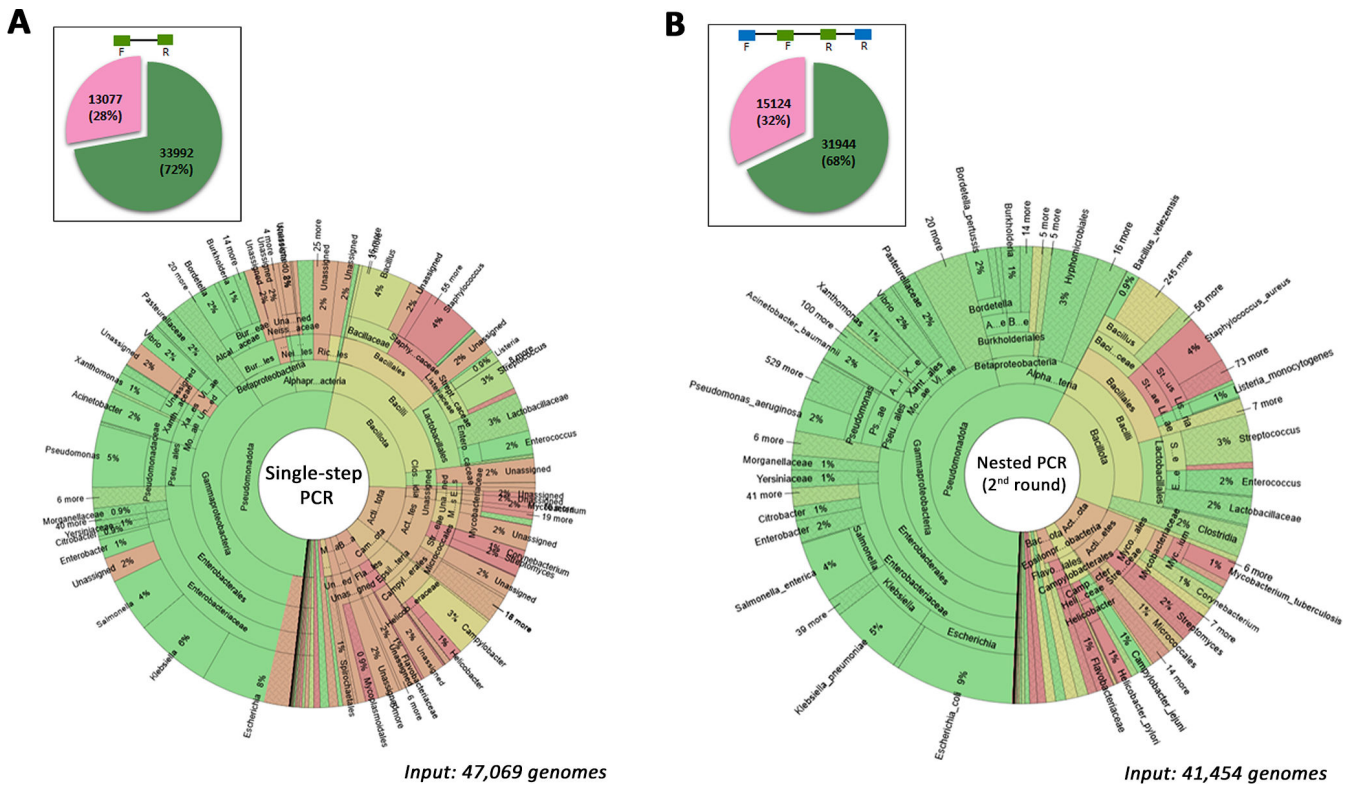
## RESULTS

### Design of single-step and nested *rpoB* PCR strategies for metabarcoding

The single-step and nested PCR strategies for the amplification of the *rpoB* gene are illustrated in Fig. 1. For the single-step PCR strategy (Fig. 1A), the region used for metabarcoding is directly amplified (35 cycles) with the *Uni\_rpoB\_deg\_F/R* inner primers containing Illumina adapters (5) to yield a 435 bp amplicon. For the nested PCR strategy (Fig. 1B), we first designed the “*rpoB\_F/R*” outer primers to bind to conserved regions of the *rpoB* gene encompassing the metabarcoding target region and generating a 906 bp amplicon. This amplicon was then subjected to a second amplification reaction with the inner primers incorporating Illumina adapters (*Uni\_rpoB\_deg\_F/R* primers), generating a 435 bp amplicon. The first PCR should enrich the substrate for the second reaction in the *rpoB* target, thereby increasing the efficiency of the second round of PCR. We optimized the number of amplification cycles using genomic DNA from a set of representative biological samples, including two mock communities, six insect larval samples, and two bacterial isolates, along with a negative control (no DNA template) (Fig. S1). Agarose gel electrophoresis showed that a total cycle number exceeding 40 resulted in non-specific amplification in the negative control. The optimal conditions—25 cycles for the first PCR and 15 for the second—yielded clear and specific amplification across biological samples, with no detectable amplicon in the negative control (Fig. S1). This optimization minimized the total number of cycles while ensuring a robust signal for Illumina sequencing.

### *In silico* evaluation of the taxonomic coverage of *rpoB* primers

We checked the universality of the outer and inner *rpoB* primers *in silico* against an *rpoB* database with 47,069 entries. This database was updated in 2024 ([https://web-genobioinfo.toulouse.inrae.fr/frogs\\_databanks/assignation/rpoB/](https://web-genobioinfo.toulouse.inrae.fr/frogs_databanks/assignation/rpoB/)). The single-step and nested strategies had the potential to amplify 72% and 68%, respectively, of the 47,069 sequences stored in the *rpoB* database (Fig. 2). A Krona representation identified the



**FIG 2** Coverage of bacterial taxonomic diversity, evaluated with the EcoPCR v.1.0.1 *in silico* tool for (A) the single-step *rpoB* PCR strategy and (B) the nested *rpoB* PCR strategy. The *in silico* PCR was performed on the 47,069 *rpoB* sequences stored in the *rpoB* database. The pie charts illustrate the number of sequences amplified with two mismatches allowed (green) and the number of unretrieved sequences (pink) from the 47,069 sequences. Below, the Krona representations display the major taxa potentially amplified by the *rpoB* marker (green) in the single-step and nested PCR strategies. Taxa that are not amplified are shown in light red. For each condition, the number of input genomes used in the *in silico* PCR simulations is indicated below the corresponding Krona plot: 47,069 genomes for the single-step PCR (panel A) and 41,454 genomes for the nested PCR (panel B), corresponding to the sequences successfully amplified during the first round (PCR1) and subsequently used as input for the second round (PCR2). For more details on the taxa covered by the two PCR strategies, all Krona visualizations are accessible via the interactive link: [https://github.com/geraldinepascal/RPOB\\_paper/tree/main/krona\\_figure](https://github.com/geraldinepascal/RPOB_paper/tree/main/krona_figure).

bacterial taxa potentially amplified by the *rpoB* primers in the single-step PCR strategy (Fig. 2A) and after the first (Fig. S2) and second rounds (Fig. 2B) of the nested PCR strategy. The outer primers used in the first PCR of the nested approach had a high degree of taxonomic coverage, encompassing over 88% of all *rpoB* sequences, accounting for the similar levels of taxonomic coverage achieved with the single-step and nested strategies. The Pseudomonadota and most of the Bacillota, excluding the *Staphylococcaceae*, were well covered by the successive use of outer *rpoB* primers in the first PCR and inner *rpoB* primers in the second PCR. By contrast, the Actinomycetota and Bacteroidota were poorly covered by both the single-step and nested PCR strategies. For more details on taxonomic coverage, an interactive link is available from [https://github.com/geraldinepascal/RPOB\\_paper/tree/main/krona\\_figure](https://github.com/geraldinepascal/RPOB_paper/tree/main/krona_figure).

**Experimental comparison of the single-step and nested *rpoB* PCR strategies with different mock samples**

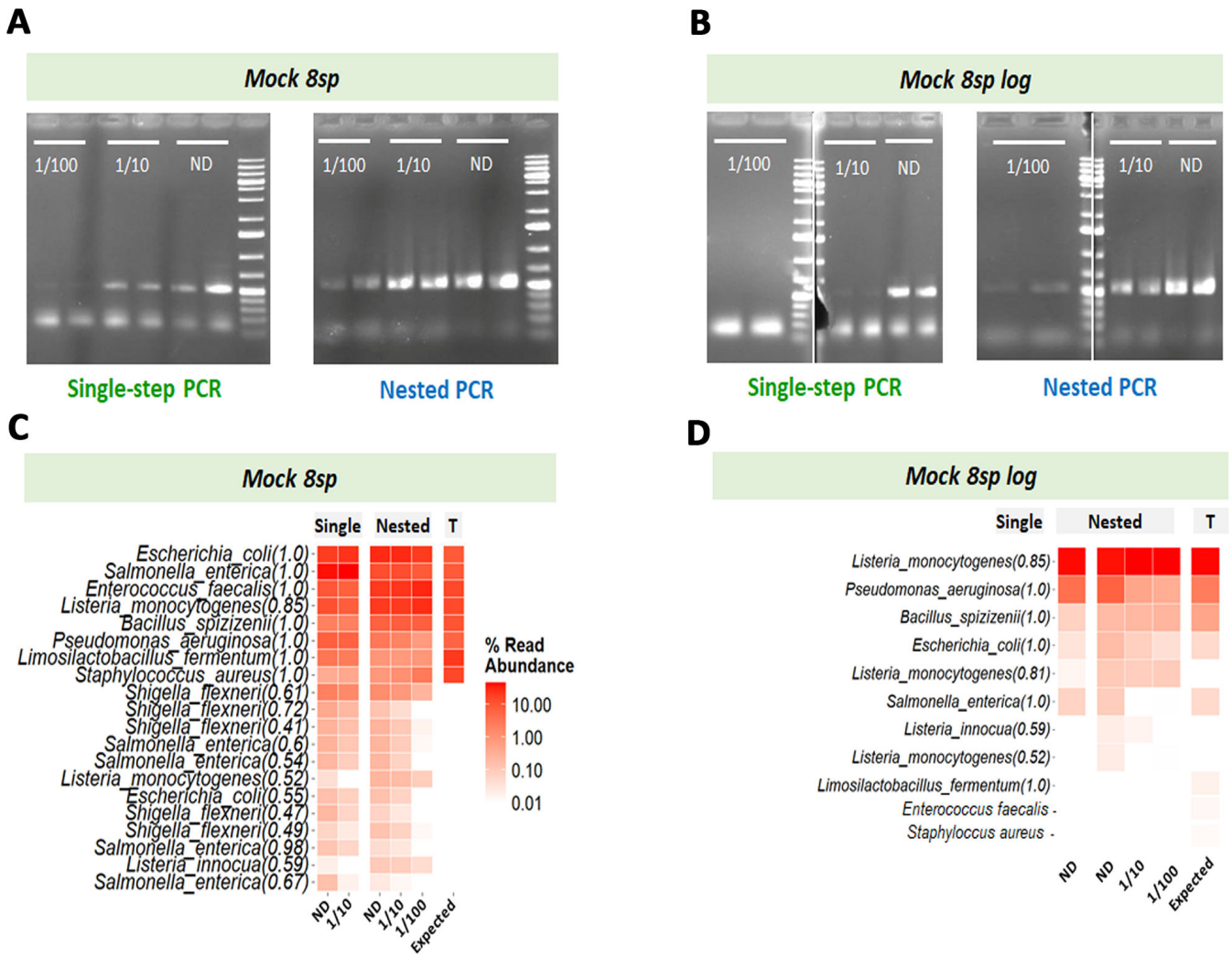
We compared the sensitivity and reliability of nested and single-step PCR strategies on two commercial mock samples—Mock\_8sp and Mock\_8sp\_log (see Table 1 for composition)—containing DNA from eight bacterial species, non-diluted (ND) or diluted (1/10 or 1/100). After amplification by PCR, gel electrophoresis of the *rpoB* amplicons showed that the nested PCR strategy resulted in successful amplification, even at dilutions as high as 1:100, for both mock samples (Fig. 3A and B). By contrast, with the single-step PCR strategy, amplification was possible only up to a dilution of 1/10 dilution

TABLE 1 Composition of the commercial DNA mock samples

Species	Expected composition (%)			
	Genomic DNA	16S copies	Genome copies	Cell numbers
Mock_8sp				
<i>Pseudomonas aeruginosa</i>	12	4.2	6.1	6.1
<i>Escherichia coli</i>	12	10.1	8.5	8.5
<i>Salmonella enterica</i>	12	10.4	8.7	8.8
<i>Limosilactobacillus fermentum</i>	12	18.4	21.6	21.9
<i>Enterococcus faecalis</i>	12	9.9	14.6	14.6
<i>Staphylococcus aureus</i>	12	15.5	15.2	15.3
<i>Listeria monocytogenes</i>	12	14.1	13.9	13.9
<i>Bacillus spizizenii</i>	12	17.4	10.3	10.3
Mock_8sp_log				
<i>Listeria monocytogenes</i>	89.1	95.9	94.8	94.9
<i>Pseudomonas aeruginosa</i>	8.9	2.8	4.2	4.2
<i>Bacillus spizizenii</i>	0.89	1.2	0.7	0.7
<i>Escherichia coli</i>	0.089	0.069	0.058	0.058
<i>Salmonella enterica</i>	0.089	0.07	0.059	0.059
<i>Limosilactobacillus fermentum</i>	0.0089	0.012	0.015	0.015
<i>Enterococcus faecalis</i>	0.00089	0.00067	0.001	0.001
<i>Staphylococcus aureus</i>	0.000089	0.0001	0.0001	0.0001

for Mock\_8sp (Fig. 3A) and only with the undiluted sample for the Mock\_8sp\_log (Fig. 3B). To ensure that the improved sensitivity of the nested PCR was not solely due to the total number of amplification cycles (25 cycles for PCR1 followed by 15 cycles for PCR2), we performed a single-step PCR using the same total number of cycles (40 cycles) to enable a direct comparison between the two approaches. As previously observed with 35 cycles, single-step PCR with 40 cycles still fails to clearly amplify the 1/100 diluted Mock\_8sp community (Panel A, Fig. S3), highlighting the advantage of the two-step amplification strategy. The nested PCR strategy was, therefore, more sensitive than the direct amplification approach. No amplification was observed from the negative control, regardless of the strategy used.

We then sequenced the amplicons visible on agarose gels with Illumina MiSeq technology. The number of reads per sample and the ASV (amplicon sequence variant) composition are detailed in Table S1; rarefaction curves and alpha-diversity analyses of the sequenced mock samples are shown in Fig. S4. The nested PCR strategy yielded alpha-diversity indices closer to the expected values than the single-step PCR, suggesting it introduces fewer spurious ASVs and reduces bias. The relative abundance of ASV by taxonomic rank clearly confirmed that the *rpoB* primers provided a high degree of taxonomic resolution, with 100% of taxa assigned to species level in the mock samples (Fig. S5). The bacterial compositions obtained for Mock\_8sp after the single-step and nested amplification strategies closely resembled the theoretical composition of the mock sample (Fig. 3C and D). However, several additional sequences were assigned to the species *Shigella flexneri* in Mock\_8sp and *Listeria innocua* in both mock samples, regardless of the strategy used. These additional assignments probably correspond to artifactual sequence variants, attributable to the close phylogenetic relationship between the genera *Shigella* and *Escherichia*, or between *L. innocua* and *L. monocytogenes*. However, neither *E. faecalis* nor *S. aureus* was detected in Mock\_8sp\_log by either single-step or nested PCR, consistent with the low proportion of genomes attributable to these species in Mock\_8sp\_log, below 0.1%, the threshold generally required for signal detection by an amplicon metabarcoding approach. We then compared the relative abundance of taxa at the species level obtained for the Mock\_8sp samples with the single-step and nested *rpoB* PCR strategies (Fig. S6). Both PCR strategies overestimated the abundances of *E. coli*, *E. faecalis*, and *L. monocytogenes*, while significantly



**FIG 3** Comparison of single-step and nested *rpoB* PCR strategies with commercial mock communities used as DNA matrices to assess amplification sensitivity (A and B) and the ability to determine taxonomic composition by metabarcoding (B and C). Panels A and B show agarose gel electrophoresis of *rpoB* amplicons from Mock\_8sp and Mock\_8sp\_log samples, respectively, amplified by single-step or nested PCR strategies. The dilutions of the mock samples are indicated on the agarose gel, below the wells. For each dilution, two technical replicates are shown. Panels C and D show heatmaps for the bacterial taxonomic composition obtained by *rpoB* metabarcoding of the Mock\_8sp and Mock\_8sp\_log samples, respectively (only samples with successful amplification were sequenced). The columns “Single” and “Nested” correspond to the single-step and nested *rpoB* PCR strategies, and the column “T” indicates the expected composition of the mocks. The dilutions of the mock samples are indicated at the bottom of the heatmap. The 20 most abundant amplicon sequence variants across samples, assigned at species level, are listed on the left. Each species name is followed by an Ribosomal Database Project (RDP) bootstrap confidence score, indicating the confidence level of the taxonomic assignment. The relative read abundance is represented by a gradient of red hues.

underestimating the abundances of *Limosilactobacillus fermentum* and *S. aureus*. This result is consistent with the *in silico* analysis of the taxonomic diversity targeted by the *rpoB* primers, which revealed limited coverage for certain bacterial groups within the Bacillota (see above). The single-step PCR strategy significantly overestimated the abundance of *Salmonella enterica*, whereas the nested PCR strategy did not (Fig. S6). To enable a robust comparison of the reproducibility and accuracy of microbial profiles generated by nested versus single-step PCR strategies, we computed Bray-Curtis dissimilarities among replicates and against the expected community composition for the Mock\_8sp samples. Analysis of multivariate dispersion revealed a lower variability among replicates in the nested PCR group compared to the single-step strategy (Fig. S7), although this difference was not statistically significant ( $F = 3.25, P = 0.113$ ). We then assessed the fidelity of each strategy by comparing Bray-Curtis distances between

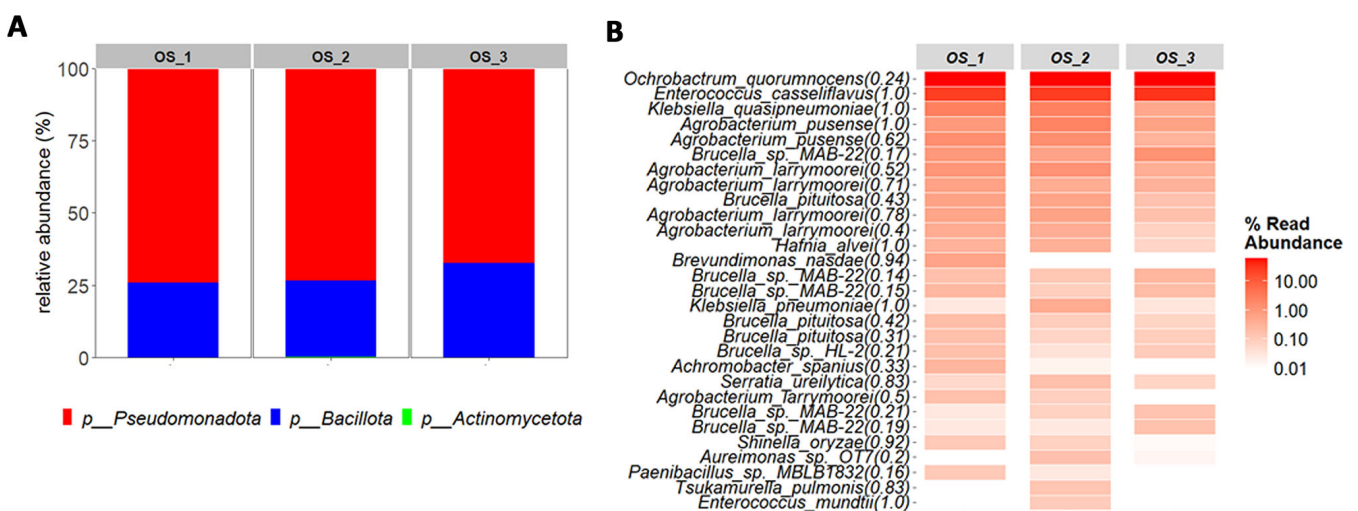
each replicate and the expected community composition. Nested PCR replicates showed significantly smaller distances to the expected profile (mean ± SD = 0.43 ± 0.00) than single-step PCR replicates (0.53 ± 0.05), indicating improved accuracy in community reconstruction (Wilcoxon test,  $P = 0.0095$ ; permutation test,  $P = 0.004$ ). These results support the higher reliability of the nested approach.

Overall, our results demonstrate that the nested PCR strategy provides taxonomic assignment profiles similar to those generated by the single-step method, but with a greater sensitivity and more efficient amplification for samples containing a low concentration of the target DNA.

### Application of the nested *rpoB* strategy to characterization of the microbiota of insect samples

We experimentally validated the nested PCR strategy on an insect, *Spodoptera frugiperda*, sampled in Senegal (near Velingara). We tested our methodology on two types of biological matrix: (i) DNA extracted from the OS of larvae, a biological fluid composed of saliva, the gut microbiota, and a food bolus (14), which contains only very low concentrations (<4 µg/mL) of bacterial DNA, and (ii) DNA extracted from half larvae (caterpillars), in which the bacterial DNA is embedded in a eukaryotic DNA matrix. The *rpoB* region was targeted by both the single-step and nested PCR strategies. PCR amplifications were performed on the six replicate samples of larval DNA (three from stage L3, three from stage L5), on DNA from the pooled OS samples, for which three PCR replicates were processed, and on the extraction control samples. The single-step *rpoB* PCR strategy generated no visible amplicon (or only faint bands) for either the OS or larval samples, after both 35 and 40 PCR cycles (panel B, Fig. S3), making them unsuitable for metabarcoding sequencing. By contrast, the nested *rpoB* strategy successfully amplified all OS and larval samples, yielding sufficient DNA for metabarcoding. Importantly, no amplification was observed for the extraction or PCR-negative controls following the two successive PCR rounds of the nested *rpoB* strategy.

All amplicons generated by the *rpoB* nested strategy were further subjected to Illumina MiSeq sequencing and statistical analysis. Details of the read count per sample and ASV composition are provided in Table S1, and rarefaction curves are provided in Fig. S8.



**FIG 4** Composition of the bacterial community according to metabarcoding of OS samples from *S. frugiperda* amplified by the nested *rpoB* PCR strategy. (A) At phylum level, the relative abundances are represented on bar plots, with sample names indicated at the top of the plots. (B) At species level, the 30 most abundant ASVs across samples are listed on the left of the heatmap. Each species name is followed by an RDP bootstrap confidence score, indicating the confidence level of the taxonomic assignment. The relative abundance (percentage) is represented by a gradient of red hues. Sample names, corresponding to the three PCR replicates, are indicated at the top of the heatmap columns.

The Pseudomonadota largely dominated in the pooled OS sample, accounting for three quarters of the observed diversity, with Bacillota accounting for the remaining quarter (Fig. 4A). Two ASVs predominated (Fig. 4B): one assigned to the genus *Ochrobactrum*, for which species-level identification was not possible due to an Ribosomal Database Project (RDP) confidence score below 0.8—the minimum threshold we considered for reliable assignment; and the other confidently identified as *Enterococcus casseliflavus*. There were slight variations between technical PCR replicates, highlighting the importance of such replicates for ensuring reliable and reproducible results for the metabarcoding of low-concentration DNA samples.

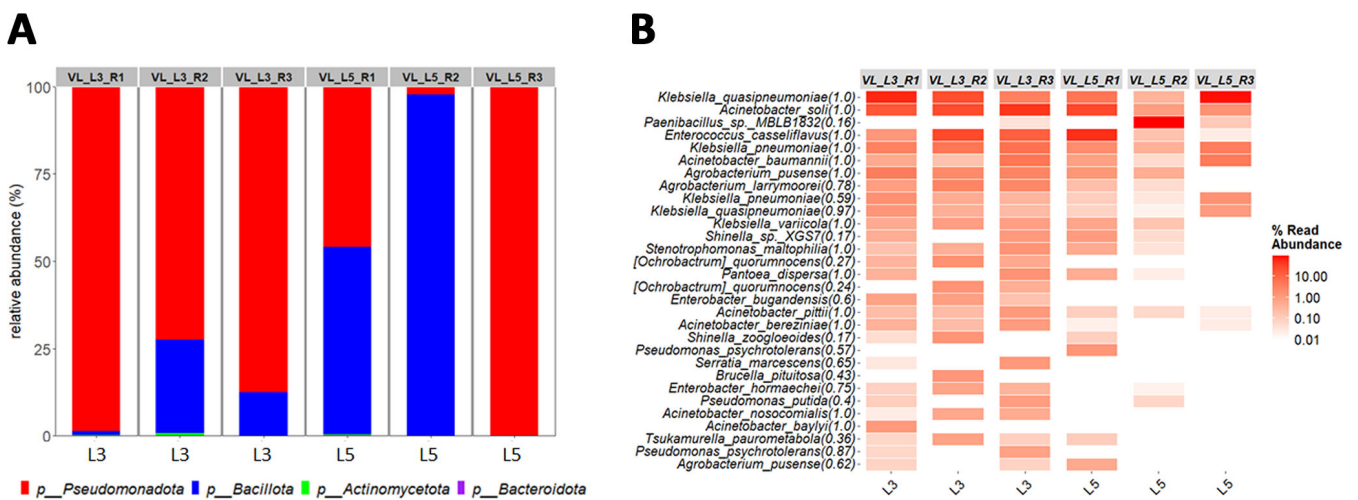
For larval samples, the nested *rpoB* strategy revealed a predominance of Pseudomonadota in five of the six biological replicates, with only one L5-stage sample containing a majority of Bacillota (Fig. 5A). The nested *rpoB* strategy enabled confident species-level identification for most ASVs (i.e., RDP assignment threshold >0.8), allowing accurate identification of the species associated with *S. frugiperda* larvae (Fig. 5B). In particular, the metabarcoding analysis revealed the prevalence of *Klebsiella quasipneumoniae*, *Acinetobacter soli*, *E. casseliflavus*, *Klebsiella pneumoniae*, and *Acinetobacter baumannii* in both L3-stage and L5-stage caterpillars. Interestingly, a few taxa, such as *E. casseliflavus* and *K. quasipneumoniae*, were common to both the larval and OS samples.

**DISCUSSION**

Metabarcoding is a powerful tool for simultaneously identifying multiple taxa within a DNA sample. However, short-read metabarcoding studies, which typically rely on amplicons of variable regions of the 16S rRNA gene to characterize bacterial microbiota (15–17), lack the precision required for fine-scale taxonomic resolution (18, 19).

In this study, we focused on the *rpoB* marker due to its ability to provide species-level taxonomic assignments (5), providing a promising alternative to 16S rRNA-based approaches.

However, the primers for *rpoB* have several weaknesses. Due to the lack of highly conserved regions in the *rpoB* gene, the primers contain multiple degenerate bases, which can make specific amplification difficult, particularly in matrices with low concentrations of bacterial gDNA and in host-associated samples, which generally contain much less prokaryotic than eukaryotic DNA. For example, in our study, even with



**FIG 5** Composition of the bacterial community revealed by metabarcoding of *S. frugiperda* demi-larva samples subjected to amplification with the nested *rpoB* PCR strategy. (A) At phylum level, relative abundances are represented on bar plots, with sample names indicated at the top of the plots. At the bottom of the graph, L3 corresponds to the third developmental stage, and L5 corresponds to the fifth developmental stage of the caterpillar (B) At species level, the 30 most abundant ASVs across samples are listed on the left of the heatmap. Each species name is followed by an RDP bootstrap confidence score, indicating the confidence level of the taxonomic assignment. The relative abundance (percentage) is represented by a gradient of red hues. Sample names are indicated at the top of the heatmap columns.

35 amplification cycles, PCR with *rpoB* primers fails to amplify the target efficiently in some complex host-associated microbiota, such as insect samples. Bivand et al. (20) tried to improve the specificity of the primers for target binding by designing *rpoB* primers according to the dual priming oligonucleotide principle and coupling the amplification with Sanger sequencing to detect pathogenic taxa in clinical samples dominated by human DNA. This approach was effective for targeted detection but is not suitable for NGS-based metabarcoding.

In this study, we validated a nested PCR strategy to improve the amplification efficiency and sensitivity of *rpoB* primers for metabarcoding analysis. Unlike the standard single-step *rpoB* PCR, this nested PCR involves a first PCR to enrich the sample in the target DNA and a second PCR to amplify the *rpoB* fragments for metabarcoding. Nested PCR has long been known to increase PCR sensitivity (21) but, to our knowledge, this is the first time that this method has been used in a metabarcoding approach.

We first demonstrate the reliability of nested PCR with the *rpoB* marker for the accurate identification of bacterial species in different mock samples of known composition and variable complexity, and its ability to increase amplification sensitivity relative to the single-step PCR strategy. Some studies have reported that nested PCR has the drawback of potentially increasing bias through preferential amplification in the two successive PCR steps (22). By optimizing the number of cycles between the first and second PCRs (40 cycles in total), we minimized the biases associated with nested PCR, ensuring robust results comparable to those obtained by single-step PCR (35 cycles in total).

We then compared the performance of the single-step and nested *rpoB* PCR strategies for describing the host-associated bacterial microbiota in two types of lepidopteran insect samples: (i) *S. frugiperda* larval OS, corresponding to a matrix containing low concentrations of bacterial DNA and, (ii) *S. frugiperda* larvae, corresponding to a eukaryotic matrix with low proportions of bacterial DNA. The single-step *rpoB* PCR failed to amplify the target, whereas the *rpoB* nested PCR strategy successfully amplified the target from all insect samples. The *rpoB* nested PCR strategy is, therefore, considerably more sensitive than the single-step *rpoB* PCR for samples with low concentrations of DNA and samples in which eukaryotic DNA predominates over bacterial DNA.

However, our findings also reveal several limitations of the *rpoB* primers, such as their limited coverage of certain bacterial phyla, contrasting with the universality of 16S primers. Indeed, the *rpoB* primers cover only about 70% of bacterial diversity (essentially the phyla Pseudomonadota and Bacillota). For example, *rpoB* lacks sensitivity for the detection of Bacteroidota, members of which may be present in host-associated microbiota samples, such as insect samples. As a means of addressing this bias, we propose the design of a multi-marker approach using phylum-specific *rpoB* primer pairs to achieve an optimal coverage of bacterial diversity.

In conclusion, in metabarcoding studies, combining a nested PCR strategy with the use of the *rpoB* marker appears to be a promising approach for describing the bacterial complexity and richness of insect-associated microbiota directly at species level. Our study highlights the benefits of nested PCR for efficiently amplifying prokaryotic DNA from eukaryotic hosts. The nested PCR *rpoB* metabarcoding method, with its species-level resolution and enhanced sensitivity, offers new opportunities for investigating the ecological roles of specific bacterial taxa in insect physiology, health, and environmental interactions. Finally, it would be of interest to extend this strategy to other host-associated microbiota samples. This study provides valuable insights for the growing scientific community exploring bacterial communities through metabarcoding across diverse ecosystems.

## MATERIALS AND METHODS

### Biological material

The two commercial mock samples contained DNA from eight bacterial strains in equal proportions (Mock\_8sp; ZymoBIOMICS Microbial Community DNA Standard; D6305) and in varying proportions (Mock\_8sp\_log; ZymoBIOMICS Microbial Community DNA Standard II\_log distribution; D6311), respectively. The composition of the mock communities is described in Table 1.

Wild *S. frugiperda* larvae and their OS were collected from maize (*Zea mays*, Poaceae) fields near Velingara in the Anambé Valley in Southern Senegal (13.163268/-14.039137) in September 2021. We collected OS from approximately 100 larvae with flexible forceps that had been disinfected with 70% ethanol and sterile plastic Pasteur pipettes. The samples were pooled to obtain a final volume of 200  $\mu$ L and stored at approximately 4°C for a few hours before being frozen at -20°C.

### Extraction of bacterial DNA from insect samples

Total DNA was extracted from both the OS pool (about 200  $\mu$ L) and six individual larvae (posterior half body): three biological replicates at stage L3 and three at stage L5. OS and larval samples were lysed by freezing at -80°C for 15 min and then heating at 80°C for 20 min. We then added 100  $\mu$ L (500  $\mu$ L for the larval samples) of Quick Extract lysis solution (Bacterial DNA extraction kit from Epicentre, USA), 1  $\mu$ L (2  $\mu$ L for the larval samples) of Ready-Lyse Lysozyme Solution (Epicentre, USA), and 20  $\mu$ L (50  $\mu$ L for the larval samples) of EDTA (0.5 M, pH 8). The samples were mechanically lysed with six 2 mm glass beads and three cycles of 7 m/s for 40 s in a FastPrep device (MP Biomedicals, Illkirch-Graffenstaden, France). The complete lysis of insect and prokaryotic cells was achieved by incubation at room temperature for 90 min and then at 55°C with 10  $\mu$ L (20  $\mu$ L for the larval samples) proteinase K (20 mg/mL), with shaking at 1,200 rpm until the solution cleared. RNA was eliminated by incubation with 10  $\mu$ L (20  $\mu$ L for the larval samples) RNase A at 20 mg/mL (Invitrogen PureLink™ RNaseA, France) for 15 min at 37°C. After a final incubation with 100  $\mu$ L protein precipitation solution (Wizard kit, Promega, France) for 5 min on ice, the samples were centrifuged for 5 min at 17,000  $\times$  g and 4°C, and the supernatant was recovered. OS samples were subjected to phenol and chloroform treatments due to the presence of polysaccharides. The DNA present in the OS and larval samples was precipitated with absolute ethanol and washed with 70% ethanol. The extracted DNA was resuspended in 50  $\mu$ L ultrapure water and stored at -20°C. We assessed the levels of contaminating bacterial DNA at each of the various steps in DNA sample preparation, by including several negative control extractions with sterile ultrapure water.

### Library preparation and Illumina MiSeq sequencing

We used two PCR strategies to generate *rpoB* amplicon libraries for MiSeq sequencing: single-step PCR, involving direct amplification, and nested PCR, consisting of two successive PCRs (Fig. 1). Each PCR was performed in duplicate (mock samples) or triplicate (OS samples). Our protocol also included RNase-free water as a negative control and positive *Enterococcus mundtii* genomic DNA as a positive control.

For the single-step *rpoB* PCR, we used the existing Univ\_ *rpoB*\_deg\_F (5' - *TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG N* GGYTWYGAAGTNCGHGACGTDCA - 3') and Univ\_ *rpoB*\_deg\_R (5' - *GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GN* T GACGYTG-CATGTTBGMRCCCATMA - 3') primers, which generate a 435 bp *rpoB* amplicon. These primers include a 5' nucleotide overhang with adapters for Illumina library indexing, indicated by the bases in italics in the primer sequences above. The conditions for the single-step PCR *rpoB* were as previously described (5). Briefly, 35 cycles of amplification were performed in a Bio-Rad thermocycler with 1 to 50 ng genomic DNA, the high-fidelity iProof DNA Polymerase (Bio-Rad), and an annealing temperature of 57°C.

For the nested *rpoB* PCR, we designed degenerate consensus outer *rpoB* primers—*rpoB*\_F (forward primer 5′ - [CAGYTDTCNCARTTYATGGAYCA](#) -3′) and *rpoB*\_R (reverse primer 5′ - [AGTTRTARCCDTYCCANGKCAT](#) - 3′)—based on ClustalW alignments (<http://multalin.toulouse.inra.fr/multalin/>) of *rpoB* gene sequences from bacterial reference genomes representing a broad taxonomic diversity of eubacteria. The binding sites of the selected primers correspond to *Escherichia coli* K12 nucleotide positions 1,528 to 2,434, resulting in the amplification of a 906-nucleotide portion of the *rpoB* gene. The specificity of the *rpoB* primers was checked *in silico* as previously described (5). The first round of amplification in the nested PCR was performed with the *rpoB*\_F/R primers and the GoTaq DNA Polymerase (Promega) in a Bio-Rad thermocycler, with 1 to 50 ng genomic DNA, in accordance with the enzyme manufacturer's instructions. Twenty-five amplification cycles were performed (annealing temperature of 50°C and annealing time of 45 s), generating an amplicon of 906 bp. The second round of amplification in the nested PCR was performed with 1 μL of the product of the first round of PCR, with the Univ\_ *rpoB*\_deg\_F/R primers incorporating Illumina adapters (see above) and the high-fidelity iProof DNA Polymerase (Bio-Rad). The same protocol was followed as for the single-step PCR (annealing temperature of 57°C and annealing time of 45 s), except that the number of amplification cycles was reduced to 15.

The presence of contaminating DNA was assessed in each PCR run by including negative controls with sterile ultrapure water as the template. The amounts of amplicon DNA and the size of the amplicons were systematically analyzed by agarose gel electrophoresis.

All Illumina-indexed amplicons of the appropriate size on agarose gels were purified, multiplexed, and sequenced at the Genseq platform (University of Montpellier, France). The amplicons were purified with magnetic beads (Clean PCR, Proteogene, France) and a new round of PCR was performed in a total volume of 18 μL (5 μL of product from the first round of PCR, 9 μL Phusion High-Fidelity PCR Master Mix, NEB, France, 2 μL index adapter I5, 2 μL index adapter I7). The cycling conditions were as follows: 95°C for 3 min, then 10 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation phase for 5 min at 72°C. We used a set of 384 index pairs based on the Integrated DNA Technologies (IDT's) unique dual index set to mark all samples and multiplex them in a single MiSeq run. The final PCR products were purified with magnetic beads, multiplexed, and sequenced in pairs on an Illumina MiSeq sequencer with the MiSeq v3 reagent kit (600 cycles; Illumina) according to the manufacturer's instructions. FastQ files were generated with `bcltofastq` at the end of the run. The quality of the raw data was assessed with a module developed by the MBB platform (Montpellier Bioinformatics Biodiversity platform, supported by the LabEx CeMEB, an ANR "Investissements d'avenir" program [ANR-10-LABX-04-01]) running the `multiqc` program (23).

### Bioinformatic processing of sequence data

The sequenced reads were analyzed with FROGS v.4.1.0 software (24). We successively used the `pre-processing`, `clustering_swarm`, `remove_chimera`, `cluster_filter`, `taxonomic_affiliation` tools with the parameters shown in Table S2 and on [https://github.com/geraldinepascal/RPOB\\_paper/blob/main/FROGS\\_analysis\\_results/readme.md](https://github.com/geraldinepascal/RPOB_paper/blob/main/FROGS_analysis_results/readme.md).

### Update of the *rpoB* database

We created an updated *rpoB* sequence database by downloading the genomes with the `genome_updater` tool ([https://github.com/pirovc/genome\\_updater](https://github.com/pirovc/genome_updater)). With this tool, we were able to create a RefSeq database (25) containing only complete genomes, complete chromosomes, and protein sequences. We obtained protein sequences from 46,711 bacteria. We then used `FetchMGS` (<https://github.com/motu-tool/fetchMGs.pl>) (26) to extract DNA and protein sequence files for *rpoB*. We obtained 47,069 *rpoB* nucleic acid sequences from 46,298 bacteria. The *rpoB* database is available from the FROGS website: [https://genoweb.toulouse.inrae.fr/frogs\\_databanks/assignment/rpoB/](https://genoweb.toulouse.inrae.fr/frogs_databanks/assignment/rpoB/).

## ***In silico* PCR analysis for evaluating the universality of *rpoB* primers**

We used OBITools suite v.1.2.11 (27) to evaluate the universality of the primers. The results were first formatted manually to make the identifiers unique and compatible with EcoPCR v.1.0.1 (28), the *in silico* PCR tool. We then used *obiconvert* to construct a database that could be used by EcoPCR. We obtained three different types of amplicons targeting *rpoB*, by performing different runs of EcoPCR using: (i) the Univ\_ *rpoB*\_deg\_F and Univ\_ *rpoB*\_deg\_R primers (see above for the primer sequences) to obtain typical metabarcoding amplicons (5), (ii) the nested PCR primers to perform two successive PCR, one with *rpoB*\_F and *rpoB*\_R (see above) and the other based on previous output sequences with the Univ\_ *rpoB*\_deg\_F and Univ\_ *rpoB*\_deg\_R primers. For comparisons of taxonomic specificity, we compared only species to which these markers (amp\_direct\_ *rpoB*, amp\_nested\_primary\_ *rpoB*, amp\_nested\_ *rpoB*) were common. In total, 42,061 genomes were sufficiently complete to contain the *rpoB* gene.

The bacterial diversity of the amplicons obtained is displayed with the krona tool (29). All these steps and scripts are detailed on GitHub: [https://github.com/geraldinepascal/RPOB\\_paper/tree/main/pcr\\_in\\_silico\\_script](https://github.com/geraldinepascal/RPOB_paper/tree/main/pcr_in_silico_script).

## **Bacterial community and statistical analyses**

ASV diversity and statistical analyses were performed with the R packages Phyloseq (30), Vegan (31), Microbiota Process (32), and Ampvis2 (33), as previously described (9). Two custom-written scripts were used to generate plots (Fig. S3) to display diversity by taxonomy rank: add\_multiaffi\_to\_abd\_table.py ([https://github.com/geraldinepascal/RPOB\\_paper/blob/main/FROGS\\_analysis\\_results/scripts/add\\_multiaffi\\_to\\_abd\\_table.py](https://github.com/geraldinepascal/RPOB_paper/blob/main/FROGS_analysis_results/scripts/add_multiaffi_to_abd_table.py)) and plot\_taxo\_ranks.py ([https://github.com/geraldinepascal/RPOB\\_paper/blob/main/FROGS\\_analysis\\_results/scripts/plot\\_taxo\\_ranks.py](https://github.com/geraldinepascal/RPOB_paper/blob/main/FROGS_analysis_results/scripts/plot_taxo_ranks.py)).

## **ACKNOWLEDGMENTS**

This work was funded by the SPE "MicroSpodoRice" grant from INRAE. Jean Mainguy's work was supported by "La Région Occitanie" and the European Union via the "Regional Platforms of Research and Innovation" call for proposals of the Occitanie Region as part of the Operational Program FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020.

We would like to thank the Insect Quarantine Platform (PIQ), a member of the Vectopole Sud network, for providing the infrastructure required for experiments on insect pests. We would also like to thank Gaëtan Clabots, Raphaël Bousquet, and Dylan Valenza for maintaining the *S. frugiperda* colonies at the DGIMI laboratory in Montpellier.

The metabarcoding data were produced with the GenSeq technical facilities at the "Institut des Sciences de l'Evolution" in Montpellier, with support from LabEx CeMEB, an ANR "Investissements d'avenir" program (ANR-10-LABX-04-01).

We would like to thank the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing assistance, computing, and storage resources.

We would like to thank Maimouna Cissoko and Nils Poulicard for their help during field sampling in Senegal and Victoria Mariotti for her help with bioinformatics analyses.

## **AUTHOR AFFILIATIONS**

<sup>1</sup>DGIMI, INRAE, Univ Montpellier, Montpellier, France

<sup>2</sup>GenPhySE, INRAE, ENVT, Université de Toulouse, Castanet Tolosan, France

<sup>3</sup>AIDA, CIRAD, Univ Montpellier, Montpellier, France

<sup>4</sup>DIADÉ, IRD, Univ Montpellier, Montpellier, France

<sup>5</sup>INRAE, BioinfOmics, GenoToul Bioinformatics facility, Université de Toulouse, Castanet-Tolosan, France

<sup>6</sup>INRAE, UR 875 MIAT, Université de Toulouse, Castanet-Tolosan, France

## AUTHOR ORCID*s*

Sophie Gaudriault  <http://orcid.org/0000-0003-2789-4959>

Jean-Claude Ogier  <http://orcid.org/0000-0002-5439-694X>

## FUNDING

Funder	Grant(s)	Author(s)
The Region Occitanie and the European Union	FEDER-FSE 2014-2020	Jean Mainguy
INRAE (France)	MicroSpodoRice	Laëtitia Leclerc

## AUTHOR CONTRIBUTIONS

Laëtitia Leclerc, Investigation, Methodology, Resources, Validation | Gabryelle Agoutin, Investigation, Methodology | Thierry Brévault, Resources | Antony Champion, Resources | Jean Mainguy, Methodology | Nicolas Nègre, Formal analysis, Resources | Sudeeptha Yainna, Resources | Géraldine Pascal, Investigation, Methodology, Validation, Visualization | Sophie Gaudriault, Conceptualization, Supervision, Validation | Jean-Claude Ogier, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization

## DATA AVAILABILITY

The sequence data supporting the findings of this study have been deposited in the European Nucleotide Archive under primary accession code [PRJEB81812](https://www.ebi.ac.uk/ena/record/PRJEB81812).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplemental material (Spectrum01417-25-s0001.pdf)**. Fig. S1 to S8.

**Table S1 (Spectrum01417-25-s0002.xlsx)**. ASV abundance tables in commercial mocks, oral secretion, and larval samples.

**Table S2 (Spectrum01417-25-s0003.docx)**. Parameters used for FROGS analysis.

## REFERENCES

- Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 43:e37. <https://doi.org/10.1093/nar/gku1341>
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1. <https://doi.org/10.1093/nar/gks808>
- Buetas E, Jordán-López M, López-Roldán A, D'Auria G, Martínez-Priego L, De Marco G, Carda-Diéguez M, Mira A. 2024. Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. *BMC Genomics* 25:310. <https://doi.org/10.1186/s12864-024-10213-5>
- Poirier S, Rué O, Peguilhan R, Coeuret G, Zagorec M, Champomier-Vergès M-C, Loux V, Chaillou S. 2018. Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using *gyrB* amplicon sequencing: a comparative analysis with 16S rDNA V3-V4 amplicon sequencing. *PLoS One* 13:e0204629. <https://doi.org/10.1371/journal.pone.0204629>
- Ogier J-C, Pagès S, Galan M, Barret M, Gaudriault S. 2019. *rpoB*, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing. *BMC Microbiol* 19:171. <https://doi.org/10.1186/s12866-019-1546-z>
- Roux S, Enault F, Bronner G, Debroas D. 2011. Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiol Ecol* 78:617–628. <https://doi.org/10.1111/j.1574-6941.2011.01190.x>
- Vos M, Quince C, Pijl AS, de Hollander M, Kowalchuk GA. 2012. A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. *PLoS One* 7:e30600. <https://doi.org/10.1371/journal.pone.0030600>
- Barret M, Briand M, Bonneau S, Préveaux A, Valière S, Bouchez O, Hunault G, Simoneau P, Jacques M-A. 2015. Emergence shapes the structure of the seed microbiota. *Appl Environ Microbiol* 81:1257–1266. <https://doi.org/10.1128/AEM.03722-14>
- Ogier J-C, Pagès S, Frayssinet M, Gaudriault S. 2020. Entomopathogenic nematode-associated microbiota: from monoxenic paradigm to pathobiome. *Microbiome* 8:25. <https://doi.org/10.1186/s40168-020-00800-5>
- Hassler HB, Probert B, Moore C, Lawson E, Jackson RW, Russell BT, Richards VP. 2022. Phylogenies of the 16S rRNA gene and its hypervariable regions lack concordance with core genome phylogenies. *Microbiome* 10:104. <https://doi.org/10.1186/s40168-022-01295-y>
- Thompson A, Bench S, Carter B, Zehr J. 2013. Coupling FACS and genomic methods for the characterization of uncultivated symbionts. *Methods Enzymol* 531:45–60. <https://doi.org/10.1016/B978-0-12-407863-5.00003-4>
- Ogier J-C, Son O, Gruss A, Tailliez P, Delacroix-Buchet A. 2002. Identification of the bacterial microflora in dairy products by temporal

- temperature gradient gel electrophoresis. *Appl Environ Microbiol* 68:3691–3701. <https://doi.org/10.1128/AEM.68.8.3691-3701.2002>
13. Galand PE, Ruscheweyh H-J, Salazar G, Hochart C, Henry N, Hume BCC, Oliveira PH, Perdereau A, Labadie K, Belsler C, et al. 2023. Diversity of the Pacific Ocean coral reef microbiome. *Nat Commun* 14:3039. <https://doi.org/10.1038/s41467-023-38500-x>
  14. Peiffer M, Felton GW. 2009. Do caterpillars secrete “oral secretions”? *J Chem Ecol* 35:326–335. <https://doi.org/10.1007/s10886-009-9604-x>
  15. Bukin Y, Galachyants Y, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. 2019. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci Data* 6:190007. <https://doi.org/10.1038/sdata.2019.7>
  16. Zemb O, Achard CS, Hamelin J, De Almeida M-L, Gabinaud B, Cauquil L, Verschuren LMG, Godon J-J. 2020. Absolute quantitation of microbes using 16S rRNA gene metabarcoding: a rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *Microbiol Open* 9:e977. <https://doi.org/10.1002/mbo3.977>
  17. Roume H, Mondot S, Saliou A, Le Fresne-Languille S, Doré J. 2023. Multicenter evaluation of gut microbiome profiling by next-generation sequencing reveals major biases in partial-length metabarcoding approach. *Sci Rep* 13:22593. <https://doi.org/10.1038/s41598-023-46062-7>
  18. Pan P, Gu Y, Sun D-L, Wu QL, Zhou N-Y. 2023. Microbial diversity biased estimation caused by intragenomic heterogeneity and interspecific conservation of 16S rRNA genes. *Appl Environ Microbiol* 89:e0210822. <https://doi.org/10.1128/aem.02108-22>
  19. Hrovat K, Dutilh BE, Medema MH, Melkonian C. 2024. Taxonomic resolution of different 16S rRNA variable regions varies strongly across plant-associated bacteria. *ISME Commun* 4:ycae034. <https://doi.org/10.1093/ismeco/ycae034>
  20. Bivand JM, Dyrhovden R, Sivertsen A, Tellevik MG, Patel R, Kommedal Ø. 2024. Broad-range amplification and sequencing of the *rpoB* gene: a novel assay for bacterial identification in clinical microbiology. *J Clin Microbiol* 62:e0026624. <https://doi.org/10.1128/jcm.00266-24>
  21. Song Y. 2005. PCR-based diagnostics for anaerobic infections. *Anaerobe* 11:79–91. <https://doi.org/10.1016/j.anaerobe.2004.10.006>
  22. Fan Z-Y, Li X-R, Mao D-P, Zhu G-F, Wang S-Y, Quan Z-X. 2009. Could nested PCR be applicable for the study of microbial diversity? *World J Microbiol Biotechnol* 25:1447–1452. <https://doi.org/10.1007/s11274-009-0033-3>
  23. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
  24. Escudé F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G. 2018. FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* 34:1287–1294. <https://doi.org/10.1093/bioinformatics/btx791>
  25. O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733–D745. <https://doi.org/10.1093/nar/gkv1189>
  26. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, Hingamp P, Alves R, Costea PI, Coelho LP, Schmidt TSB, Almeida A, Mitchell AL, Finn RD, Huerta-Cepas J, Bork P, Zeller G, Sunagawa S. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* 10:1014. <https://doi.org/10.1038/s41467-019-0884-4>
  27. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. 2016. obitools: a unix-inspired software package for DNA metabarcoding. *Mol Ecol Resour* 16:176–182. <https://doi.org/10.1111/1755-0998.12428>
  28. Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessièrè J, Taberlet P, Pompanon F. 2010. An *in silico* approach for the evaluation of DNA barcodes. *BMC Genomics* 11:434. <https://doi.org/10.1186/1471-2164-11-434>
  29. Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a web browser. *BMC Bioinformatics* 12:385. <https://doi.org/10.1186/1471-2105-12-385>
  30. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
  31. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
  32. Xu S, Zhan L, Tang W, Wang Q, Dai Z, Zhou L, Feng T, Chen M, Wu T, Hu E, Yu G. 2023. MicrobiotaProcess: a comprehensive R package for deep mining microbiome. *Innovation (Camb)* 4:100388. <https://doi.org/10.1016/j.xinn.2023.100388>
  33. Andersen KS, Kirkegaard RH, Karst SM, Albertsen M. 2018. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv*. <https://doi.org/10.1101/299537>