

RESEARCH ARTICLE

Open Access



# Decontamination of 16S rRNA gene amplicon sequence datasets based on bacterial load assessment by qPCR

Vladimir Lazarevic\*, Nadia Gaia, Myriam Girard and Jacques Schrenzel

## Abstract

**Background:** Identification of unexpected taxa in 16S rRNA surveys of low-density microbiota, diluted mock communities and cultures demonstrated that a variable fraction of sequence reads originated from exogenous DNA. The sources of these contaminants are reagents used in DNA extraction, PCR, and next-generation sequencing library preparation, and human (skin, oral and respiratory) microbiota from the investigators.

**Results:** For *in silico* removal of reagent contaminants, a pipeline was used which combines the relative abundance of operational taxonomic units (OTUs) in V3–4 16S rRNA gene amplicon datasets with bacterial DNA quantification based on qPCR targeting of the V3 segment of the 16S rRNA gene. Serially diluted cultures of *Escherichia coli* and *Staphylococcus aureus* were used for 16S rDNA profiling, and DNA from each of these species was used as a qPCR standard. OTUs assigned to *Escherichia* or *Staphylococcus* were virtually unaffected by the decontamination procedure, whereas OTUs from *Pseudomonas*, which is a major reagent contaminant, were completely or nearly completely removed. The decontamination procedure also attenuated the trend of increase in OTU richness in serially diluted cultures.

**Conclusions:** Removal of contaminant sequences derived from reagents based on use of qPCR data may improve taxonomic representation in samples with low DNA concentration. Using the described pipeline, OTUs derived from cross-contamination of negative extraction controls were not recognized as contaminants and not removed from the sample dataset.

**Keywords:** Contaminant DNA, Bacterial communities, 16S rRNA gene sequencing, Microbiome

## Background

The development of PCR and next-generation sequencing techniques has facilitated studying microbial communities without it being necessary to culture individual members. Because growth requirements vary greatly among different species, and for some species growth conditions have not yet been determined, molecular methods in microbiota investigations are advantageous. However, culture-free approaches may introduce biases in the experimental pipeline, starting from DNA extraction through the generation of sequencing libraries to data analysis. Identification of unexpected taxa in datasets derived from low-density microbiota [1–3], diluted

mock communities [4], and cultures [5], demonstrated that a variable fraction of sequence reads originated from exogenous DNA. The sources of these contaminants are reagents used in DNA extraction, PCR, and next-generation sequencing library preparation, and possibly human (skin, oral, and respiratory) microbiota from the investigators [6].

Sample datasets can be decontaminated by removing sequence reads assigned to operational taxonomic units (OTUs) found in negative extraction controls (NECs). Bioinformatics pipelines for performing microbiome analysis such as Qiime [7] facilitate performing this step in an automated manner. Some OTUs identified as contaminants across different studies were repeatedly assigned to the same species or genera [5]. However, an OTU that corresponds to the genuine member of the microbiota of interest may also be found in relevant

\* Correspondence: vladimir.lazarevic@genomic.ch  
Genomic Research Laboratory, Department of Genetics and Laboratory Medicine, Department of Medical Specialties, Geneva University Hospitals, Rue Gabrielle-Perret-Gentil 4, CH-1211, Geneva 14, Switzerland

NECs. It has been suggested not to remove OTUs identified in NECs if they are biologically expected in the given sample type [5]. The distinction between expected and unexpected OTUs in a given sample type may not be always straightforward. For example, *Propionibacterium* is a known reagent contaminant but it is genuinely present in the skin microbiota in proportions that vary between individuals [3]. Similarly, *Stenotrophomonas*, another common reagent contaminant, emerged as a new airway pathogen [4, 8], that may complicate the analysis of respiratory tract samples.

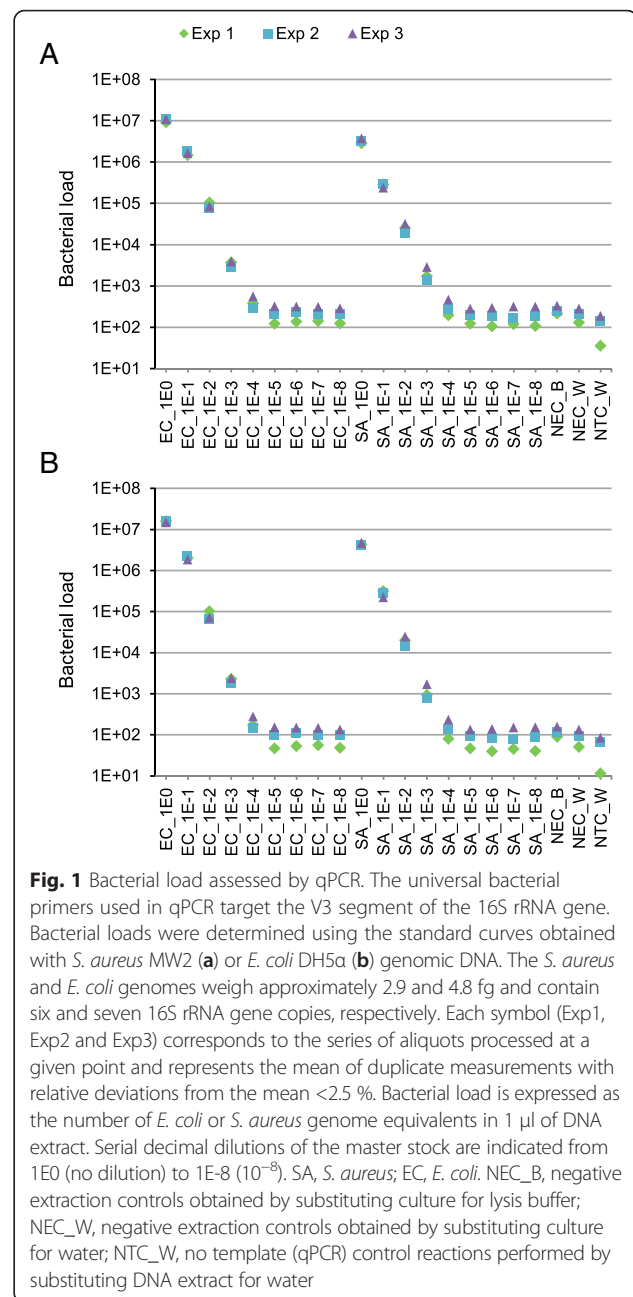
The need to recognize as many contaminants as possible based on differences in the relative abundance of bacterial taxa between NECs, low-density samples, and high-density samples has been highlighted [5, 9]. Others proposed that contaminant OTUs excluded should be those whose relative abundance in NECs is above a given threshold [10]. Inverse correlation of a taxon relative abundance with bacterial load as an indicator of a possible reagent contaminant was initially described for bacterial genera in a mock community [4] and subsequently confirmed at the OTU level in cultures and ‘real’ microbiota samples [2, 3, 5, 11].

The removal of OTUs whose mean relative abundance in NECs is higher than that of microbiota samples of interest has been used to decontaminate datasets obtained by 16S rRNA gene amplicon sequencing of relatively low-density skin and respiratory tract bacterial communities [2, 3]. However, the absolute abundance of certain OTUs may be substantially higher in microbiota samples than in relevant NECs, even if their relative abundance shows the opposite pattern. It is advisable not to remove such OTUs, as they correspond to the microbiota of interest. Here, we further develop this approach by combining relative abundance of OTUs with bacterial load in DNA extracts assessed by quantitative real-time PCR (qPCR).

## Results

### Bacterial load determined by culture and qPCR

*Staphylococcus aureus* and *Escherichia coli* overnight cultures were washed and concentrated, and resulted in  $3.5 \times 10^{10}$  and  $5.3 \times 10^9$  colony-forming units (CFU)/ml, respectively. Serial decimal dilutions of these master stocks were aliquoted in triplicate and frozen. DNA was extracted from each of the three identical series of aliquots on separate occasions. Serial culture dilutions down to  $10^{-5}$  correlated with decreasing DNA yields in purified extracts determined by qPCR targeting the V3 segment of the bacterial 16S rRNA gene with universal bacterial primers (Fig. 1). Further dilutions ( $10^{-6}$ – $10^{-8}$ ) of the master stocks had DNA quantity estimates similar to those of NECs obtained by substituting culture for water (NEC\_W) or lysis buffer (NEC\_B) in DNA



extraction. The lowest DNA concentration was found for no-template controls (NTC\_W) in which water was used instead of DNA extract. Bacterial loads determined by qPCR based on *S. aureus* (Fig. 1a) or *E. coli* (Fig. 1b) reference curves showed similar patterns.

### Taxonomic analysis of samples and negative controls

The sequence dataset generated by Illumina sequencing of V3–4 16S rRNA gene amplicons was represented by 9042–176,345 reads per sample after quality filtering and OTU mapping. The proportion of quality-filtered

sequences with no hits (with  $\geq 97$  % identity) in the Greengenes reference database [12] was  $0.63 \pm 0.04$  % (mean  $\pm$  SD) for master culture stocks and  $4.2 \pm 1.9$  % for negative controls (NEC\_W, NEC\_B, and NTC\_W). The RDP classifier [13] (with  $\geq 80$  % confidence) assigned these sequences to Pseudomonadales (24.1 %), Parcubacteria (11.7 %), Actinomycetales (8.4 %), unclassified Bacteria (14.6 %) and unclassified organisms (7.8 %).

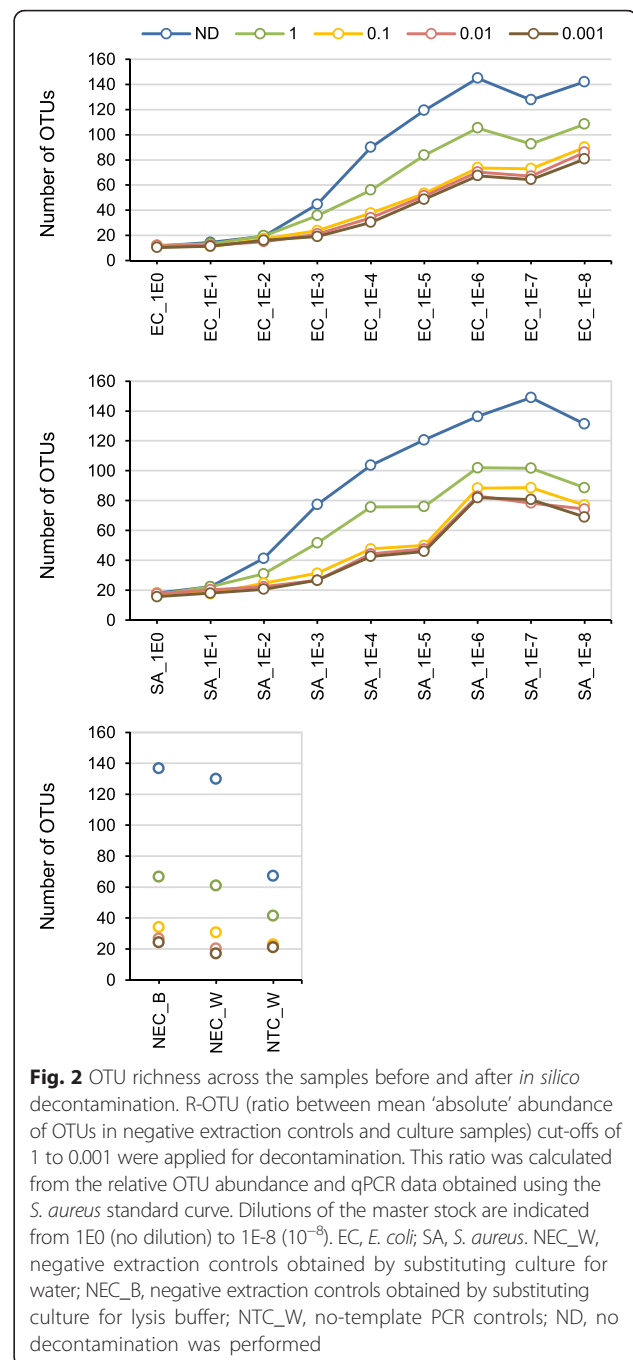
A total of 2673 OTUs were identified in the final dataset of which 1718 were found only in samples, 276 were specific to negative controls, and 855 were found in both. OTUs richness in NEC\_W and NEC\_B were similar to each other but higher than that of NTC\_W. In the dataset normalized to the same number of sequences per sample (3500), OTU richness increased as bacterial counts decreased (Fig. 2).

Serial culture dilutions were associated with a decrease in the proportion of sequence reads assigned to *Escherichia* or *Staphylococcus*, and an increase in the relative abundance of reads that derived from contaminants. *Pseudomonas* was major contaminant, and was most abundant in negative controls and highly diluted samples (Fig. 3). Interestingly, OTUs assigned to *Pseudomonas* had different profiles in NECs (NEC\_W and NEC\_B) and NTC\_W. For example, OTU4028110 and OTU1566691, which dominated NECs and NTC\_W, respectively, differed in 11 residues in the sequenced V3–4 region. These results show specific contamination of both DNA extraction and PCR reagents. The most abundant OTUs in NTC\_W, which corresponded to the contaminants of the PCR reagents, were identified in most NECs, where they were outnumbered by OTUs from contaminants from DNA extraction. However, the OTUs highly abundant in NECs were mostly absent from NTC\_W. The similarity of NEC\_W to NEC\_B and its difference from NTC\_W OTU profile indicate that ultrapure water was not the major source of DNA contamination.

In the sequence data from highly diluted *E. coli* cultures, we identified a substantial proportion of OTUs assigned to *Staphylococcus* and, also found *Escherichia* OTUs in the sequence data from diluted *S. aureus*. These results and the fact that *Staphylococcus* has not been previously recognized as reagent contaminant [5] indicate cross-contamination during DNA extraction from samples with high bacterial load, notably by *E. coli* in experiment 1 and by *S. aureus* in experiment 3 (Fig. 3). However, the proportion of cross-contaminants was lower than that of reagent contaminants.

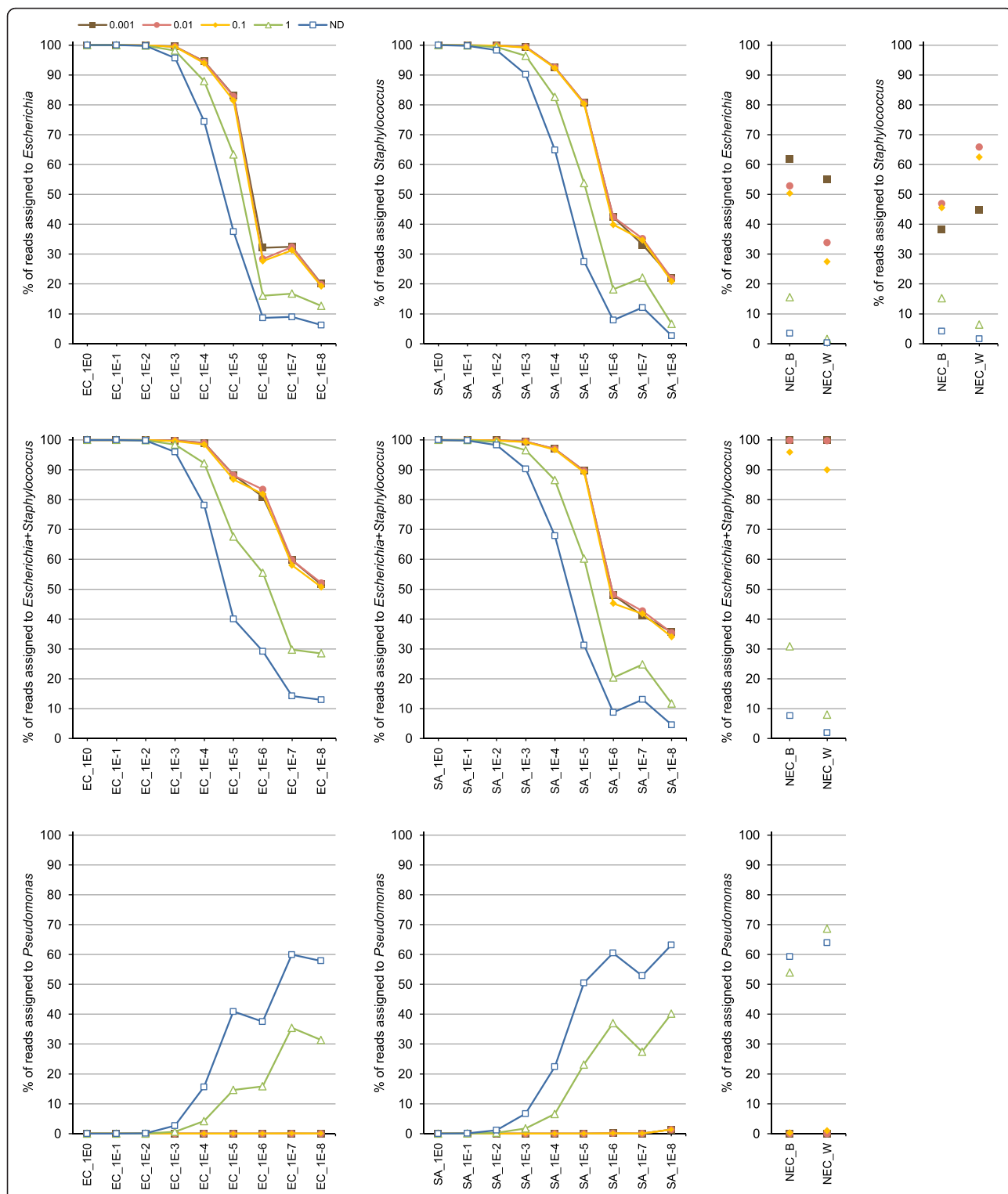
#### In silico decontamination procedure

To obtain an approximate estimation of the 'absolute' abundance of OTUs, expressed in arbitrary units, we multiplied the relative abundance of each OTU by the 16S rRNA gene copy number of a given sample



(determined by qPCR). We then calculated the ratio (designated R-OTU) between mean 'absolute' abundance of OTUs in NECs and culture samples. The dataset was decontaminated *in silico* using four R-OTU cut-off values (1, 0.1, 0.01 and 0.001), by removing OTUs for which this ratio was exceeded. The proportion of the bacterial genera *Escherichia*, *Staphylococcus* and *Pseudomonas* before and after decontamination are shown in Fig. 4. The decontamination procedure improved the taxonomic





**Fig. 4** Effect of *in silico* decontamination on taxonomic profiles of culture dilutions and negative extraction controls. Means for three samples obtained in separate DNA extraction experiments are given. The R-OTU (the ratio between mean ‘absolute’ abundance of OTUs in negative extraction controls and culture samples) cut-offs of 1 to 0.001 were applied for decontamination. This ratio was calculated from the relative OTU abundance and qPCR data obtained using the *S. aureus* standard curve. Dilutions of the master stock are indicated from 1E0 (no dilution) to 1E-8 ( $10^{-8}$ ). EC, *E. coli*; SA, *S. aureus*. NEC\_W, negative extraction controls obtained by substituting culture for water; NEC\_B, negative extraction controls obtained by substituting culture for lysis buffer; ND, no decontamination was performed

available in limited amount and/or have low bacterial load. Longer 16S rRNA gene sequences, lower sequencing error rates, and OTU clustering at identity thresholds >97 % may contribute to better distinction between contaminants and 'real' OTUs. In addition to further development of bioinformatics and statistical approaches for decontamination after sequencing, it may also be advantageous to reduce DNA contamination of laboratory reagents during their manufacturing and reduce the risk of sample-to-sample contamination during the experiments.

## Conclusions

We show that removal of contaminant OTUs derived from reagents based on the combination of qPCR data and relative abundance of OTUs may improve taxonomic representation in samples with DNA concentrations close to those of NECs. Using the described approach, OTUs derived from cross-contamination, in contrast to those derived from reagents, were not recognized as contaminants and not removed from the dataset. The approach we used in this study may prove useful in situations where OTUs identified in negative controls have higher relative abundance but lower absolute abundance compared with microbiota samples.

## Methods

### Bacterial strains and culture

Fresh colonies of *E. coli* DH5 $\alpha$  (Invitrogen, Carlsbad, CA, USA) and *S. aureus* MW2 (strain NRS 123 obtained from the Network of Antibiotic Resistance in *S. aureus* (NARSA)) were inoculated in 20 ml Difco Mueller-Hinton broth medium (BD Diagnostics, Sparks, MD, USA) and incubated overnight with shaking (180 rpm). Three overnight cultures of the same strain were pooled, centrifuged at 1600 *g* for 10 min and washed twice with NaCl 0.9 %. The cells were suspended in 6 ml ddH<sub>2</sub>O (Sigma-Aldrich, Munich, Germany). From each suspension, four serial dilutions 10<sup>-1</sup> to 10<sup>-8</sup> were performed by adding 100  $\mu$ l inoculum to 900  $\mu$ l ddH<sub>2</sub>O. One series was used immediately for plating onto Mueller-Hinton Agar (BD Diagnostics). CFU were counted after 24-h incubation at 37 °C. The other three series were placed at -20 °C and used for DNA extraction within the following 6 d.

### DNA extraction

DNA was extracted using the NucleoSpin Soil kit (Macherey-Nagel, Düren, Germany). Five hundred microliters of bacterial cell suspensions, 700  $\mu$ l of lysis buffer SL1, and 100  $\mu$ l of Enhancer SX were shaken in a NucleoSpin Bead Tube for 4 min at maximum speed on a Vortex-Genie 2 with a horizontal tube holder (Scientific Industries, New York, USA). The lysate was

centrifuged at 11,000 *g* for 1 min. Then, we followed the NucleoSpin Soil kit booklet protocol (November 2011/ Rev. 03). DNA was eluted in 50  $\mu$ l of elution buffer SE. Purified DNA was stored at -20 °C.

One *E. coli* and one *S. aureus* dilution series of samples (10<sup>0</sup>–10<sup>-8</sup>) were processed in parallel on three different days. In each batch, an NEC was performed using 500  $\mu$ l SL1 buffer (NEC\_B) or 500  $\mu$ l ddH<sub>2</sub>O (NEC\_W) instead of bacterial suspensions.

### PCR and sequencing

The V3–4 region of the bacterial 16S rRNA genes (*E. coli* positions 341–805) was amplified using template DNA from *E. coli* and *S. aureus* cultures, and from NECs. PCR was performed in a 25  $\mu$ l volume that contained 5  $\mu$ l of DNA extract, 12.5  $\mu$ l KAPA2G Robust HotStart ReadyMix (Kapa Biosystems, Boston, MA, USA), 6.5 ddH<sub>2</sub>O, and 0.5  $\mu$ l each of 10  $\mu$ M forward primer 341 F 5'-CCT ACGGGNGGCWGCAG-3' and reverse primer 805R 5'-GACTACHVGGGTATCTAATCC-3' [14]. Two NTC\_W were performed in parallel for each series of bacterial suspensions and NECs using 5  $\mu$ l ddH<sub>2</sub>O instead of DNA extract. The PCR conditions included an initial denaturation at 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 51 °C for 30 s, and extension at 72 °C for 30 s, with a final extension at 72 °C for 5 min. Each PCR was performed in duplicate and the products were combined. The pooled sample was run on a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) for quality analysis. The primers from the first round of PCR were removed by digesting 5- $\mu$ l samples with 1 unit Exonuclease I (New England Biolabs, Ipswich, MA, USA) in a total volume of 10  $\mu$ l Exonuclease I Reaction Buffer (New England Biolabs) at 37 °C for 30 min. The enzyme was inactivated at 95 °C for 15 min. Amplicon barcoding was performed by re-amplification using 1  $\mu$ l of Exonuclease I-treated first-round PCR, 15 pmol each of forward primer 5'-NNNNNNNNNNTCCTACGGGNGGCWGCAG-3' and reverse primer 5'-NNNNNNNNNNTGACTACHVGGGTATCTAAKCC-3' in a 20- $\mu$ l volume of MyTaq buffer that contained 1.5 units MyTaq DNA polymerase (Bioline, London, UK) and 2  $\mu$ l of BioStab PCR optimizer (II) (Sigma-Aldrich, Munich, Germany). For each sample, the forward and reverse primers had the same 10-nt barcode sequence. PCRs were carried out using the following parameters: pre-denaturation for 2 min at 96 °C, followed by eight cycles of 96 °C for 15 s, 50 °C for 30 s, and 70 °C for 90 s. DNA concentration of amplicons of interest was determined by gel electrophoresis. About 20 ng amplicon DNA of each sample were pooled and purified with one volume Agencourt AMPure XP beads (Beckman Coulter, Nyon, Switzerland) to remove primer dimers and other small mispriming products, followed by an additional purification using a

MinElute PCR Purification Kit (Qiagen, Venlo, the Netherlands). About 100 ng of the pooled amplicon DNA was used to construct a sequencing library using the Ovation Rapid DR Multiplex System 1–96 (NuGEN, San Carlos, CA, USA). The library was size-selected by gel electrophoresis and sequenced from both ends for 300 cycles on the Illumina MiSeq using MiSeq v3 Reagent Kit (Illumina, San Diego, CA, USA) at LGC Genomics (Berlin, Germany). Demultiplexed FASTQ files were generated from base-calls using Illumina's bcl2fastq v1.8.4 software. Reads with incorrect barcodes, missing barcodes, or conflicting barcode pairs were discarded. A maximum of three mismatches per primer were allowed. After removal of primer sequences using proprietary LGC Genomics software, forward and reverse-complemented reverse reads were merged using BBMerge from the BBSMap\_34.48 package (<http://sourceforge.net/projects/bbmap/>) with minimum overlap of 12 bases and a maximum of 3 mismatches.

#### Sequence analysis

Sequence filtering was performed using the command trim.seq in MOTHUR v1.35 [15]. The reads that contained ambiguous bases or homopolymer runs longer than 12 bases were removed. Then, sequences were truncated at the beginning of a 20-base window with an average Phred quality <30. Sequences that, after trimming, had a length <300 bases were discarded.

Denosing and clustering of 16S rDNA sequences were made by OTU mapping with the Greengenes reference database [12] pre-clustered at 97 % identity (Greengenes file 97\_otus.fasta as of 17 May 2013) using USEARCH (-usearch\_global -wordlength 30 -id 0.97 -query\_cov 0.9 -top\_hits\_only) [16]. Sequences with no hits were discarded. For sequences with multiple best hits, the hit that corresponded to the Greengenes reference sequence most frequently assigned in the entire dataset was retained. The reads were classified using naïve Bayesian method and the RDP reference database [13] via MOTHUR (command classify.seqs with options -method = wang and -cutoff = 80) and MOTHUR files trainset10\_082014.rdp.fasta and trainset10\_082014.rdp.tax. The consensus taxonomy of an OTU was defined as the taxonomy that represented most of the reads within this OTU.

#### qPCR

qPCR assay was performed on an Mx3005P qPCR system (Agilent Technologies, Santa Clara, CA, USA). Reaction mixtures contained 12.5 µl of 2× Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies), 0.75 µl of 1/250 diluted reference dye (Agilent Technologies), 0.3 µl of each 25 µM forward (5'-ACTCCTACGGGAGG CAGCAGT-3') and reverse (5'- ATTACCGCGGCTGC

TGGC-3') primers [17], 1 µl of DNA extract, and 10.15 µl water. The primers used amplify the V3 region of bacterial 16S rRNA genes (*E. coli* positions 338–534). The cycling conditions included initial denaturation of 10 min at 95 °C followed by 40 cycles of 95 °C for 30 s and 68° for 1 min. No-template qPCR controls were performed using 1 µl ddH<sub>2</sub>O instead of DNA extract. The reference curves for DNA quantitation were obtained using known concentrations of genomic DNA of *E. coli* strain DH5α and *S. aureus* strain MW2. All reactions were carried out in duplicate.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Availability of data and materials

The dataset supporting the conclusions of this article is available in the MG-RAST repository [18] under the project ID 14505 (4647144.3–4647209.3).

#### Additional files

**Additional file 1: Figure S1.** Relative abundance of bacterial genera before and after the decontamination procedure. Genera with mean relative abundance >0.5 % in negative extraction controls are presented. The proportion is indicated by the scale at the bottom of the plot. The R-OTU (ratio between mean 'absolute' abundance of OTUs in negative extraction controls and culture samples) cut-off of 0.01 was applied for decontamination. This ratio was calculated from the relative OTU abundance and qPCR data obtained using the *S. aureus* standard curve. For a given culture/dilution or negative extraction control, the data obtained from DNA extractions performed at three different time points (Exp1–Exp3) are presented from left to right. Dilutions of the master stock are indicated from 1E0 (no dilution) to 1E-8 (10<sup>-8</sup>). NEC\_W, negative extraction controls obtained by substituting culture for water; NEC\_B, negative extraction controls obtained by substituting culture for lysis buffer. (PDF 13 kb)

**Additional file 2: Figure S2.** Effect of *in silico* decontamination on taxonomic profiles of culture dilutions and negative extraction controls. Means for three samples obtained in separate DNA extraction experiments are given. The R-OTU (ratio between mean 'absolute' abundance of OTUs in negative extraction controls and culture samples) cut-offs of 1 to 0.001 were applied for decontamination. This ratio was calculated from the relative OTU abundance and qPCR data obtained using the *E. coli* standard curve. Dilutions of the master stock are indicated from 1E0 (no dilution) to 1E-8 (10<sup>-8</sup>). EC, *E. coli*; SA, *S. aureus*. NEC\_W, negative extraction controls obtained by substituting culture for water; NEC\_B, negative extraction controls obtained by substituting culture for lysis buffer; ND, no decontamination was performed. (PDF 16 kb)

#### Abbreviations

NEC: negative extraction control; NEC\_B: negative extraction control obtained by substituting culture for lysis buffer; NEC\_W: negative extraction control obtained by substituting culture for water; NTC\_W: no-template PCR (or qPCR) control; OTU: operational taxonomic unit; R-OTU: the ratio between mean 'absolute' abundance of OTUs in NEC and culture samples.

#### Competing interests

The authors declare that they have no competing interests.

**Authors' contributions**

VL and JS designed the experiments and drafted the manuscript. MG performed the experiments. NG and VL performed the bioinformatics and taxonomic analyses. All authors approved the final manuscript.

**Acknowledgements**

We thank NARSA for providing the strain MW2 and Stefano Leo for assistance with manuscript revision.

**Funding**

No funding was received for this study.

Received: 24 August 2015 Accepted: 14 April 2016

Published online: 23 April 2016

**References**

- Biesbroek G, Sanders EA, Roeselers G, Wang X, Caspers MP, Trzcinski K, et al. Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS ONE*. 2012;7:e32942.
- Lazarevic V, Gaia N, Emonet S, Girard M, Renzi G, Despres L, et al. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol*. 2014;4:65.
- Troccez M, Gaia N, Beccucci S, Schrenzel J, Cayeux I, Starckenmann C, et al. Mapping axillary microbiota responsible for body odours using a culture-independent approach. *Microbiome*. 2015;3:3.
- Willner D, Daly J, Whiley D, Grimwood K, Wainwright CE, Hugenholtz P. Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS ONE*. 2012;7:e34605.
- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
- Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. *Nat Methods*. 2011;8:761–3.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
- Surette MG. The cystic fibrosis lung microbiome. *Ann Am Thorac Soc*. 2014;11 Suppl 1:S61–5.
- Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of experimental contamination in microbiome studies. *Genome Biol*. 2014;15:564.
- Flores GE, Henley JB, Fierer N. A direct PCR approach to accelerate analyses of human-associated microbial communities. *PLoS ONE*. 2012;7:e44563.
- Jervis-Bardy J, Leong LE, Marri S, Smith RJ, Choo JM, Smith-Vaughan HC, et al. Deriving accurate microbiota profiles from human samples with low bacterial content through post-sequencing processing of Illumina MiSeq data. *Microbiome*. 2015;3:19.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J*. 2012;6:610–8.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42:D633–42.
- Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J*. 2011;5:1571–9.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537–41.
- Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
- Hartman AL, Lough DM, Barupal DK, Fiehn O, Fishbein T, Zasloff M, et al. Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc Natl Acad Sci U S A*. 2009;106:17187–92.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008;9:386.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

