High-throughput qPCR and 16S rRNA gene amplicon sequencing as complementary methods for the investigation of the cheese microbiota

Matthias Dreier[1,2*], Marco Meola[1,3,4,5], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1] and Pilar Junier[2]

Abstract

Background: Next-generation sequencing (NGS) methods and especially 16S rRNA gene amplicon sequencing have become indispensable tools in microbial ecology. While they have opened up new possibilities for studying microbial communities, they also have one drawback, namely providing only relative abundances and thus compositional data. Quantitative PCR (qPCR) has been used for years for the quantification of bacteria. However, this method requires the development of specific primers and has a low throughput. The constraint of low throughput has recently been overcome by the development of high-throughput qPCR (HT-qPCR), which allows for the simultaneous detection of the most prevalent bacteria in moderately complex systems, such as cheese and other fermented dairy foods. In the present study, the performance of the two approaches, NGS and HT-qPCR, was compared by analyzing the same DNA samples from 21 Raclette du Valais protected designation of origin (PDO) cheeses. Based on the results obtained, the differences, accuracy, and usefulness of the two approaches were studied in detail.

Results: The results obtained using NGS (non-targeted) and HT-qPCR (targeted) show considerable agreement in determining the microbial composition of the cheese DNA samples studied, albeit the fundamentally different nature of these two approaches. A few inconsistencies in species detection were observed, particularly for less abundant ones. The detailed comparison of the results for 15 bacterial species/groups measured by both methods revealed a considerable bias for certain bacterial species in the measurements of the amplicon sequencing approach. We identified as probable origin to this PCR bias due to primer mismatches, variations in the number of copies for the 16S rRNA gene, and bias introduced in the bioinformatics analysis.

Conclusion: As the normalized microbial composition results of NGS and HT-qPCR agreed for most of the 21 cheese samples analyzed, both methods can be considered as complementary and reliable for studying the microbial composition of cheese. Their combined application proved to be very helpful in identifying potential biases and overcoming methodological limitations in the quantitative analysis of the cheese microbiota.

Introduction

Molecular biology methods are increasingly replacing classical microbiological methods for the investigation of microbial communities in food products. More specifically, new developments in sequencing technology have made the use of next-generation sequencing (NGS) more affordable and widely applicable. Especially, 16S rRNA

*Correspondence: matthias.dreier@agroscope.admin.ch
[2] Laboratory of Microbiology, University of Neuchâtel, Emile-Argand 11, CH-2000 Neuchâtel, Switzerland
Full list of author information is available at the end of the article

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 2 of 18

gene amplicon sequencing is becoming increasingly widespread to investigate the composition of bacterial communities in a variety of ecosystems. Nevertheless, the optimization and validation of such approaches can be a challenging task, especially because many different aspects have to be considered to achieve reliable results [1, 2]. For instance, the comparison of commonly used sequencing platforms, databases, and classification algorithms applied to mock communities consisting of bacterial species found in dairy products has revealed significant variations in the observed bacterial community compositions [3]. Furthermore, the accurate taxonomic classification of species in complex samples remains a challenging task, which depends on many factors, such as the selected primers for the variable 16S rRNA gene region, the taxonomy assignment method, and the database used [4–7]. Ecosystem-based databases for taxonomy assignment can achieve higher resolution at the species level [8–11] as shown by an improvement in species level classification obtained with a specific and manually curated database for milk and cheese analysis as compared to more general databases [9].

Apart from the comparison of the sequencing platforms and the bioinformatics analysis, it is relatively complex to validate the results of NGS with other approaches, as most other methods do not provide comparable in-depth data. For instance, a study comparing traditional culture methods and NGS in fecal and hypopharyngeal samples of healthy children found that the second method identified 7 to 20 times more unique species [12]. Most frequently, the 16S rRNA gene amplicon and shotgun sequencing methods have been compared to each other [13–17]. However, a large study of microbial communities in lakes in Brazil has reported a weak correlation and major differences in taxonomic diversity and abundance data between the two methods, with amplicon sequencing detecting significantly more phylum- and family-level diversity [17].

An inherent limitation of the amplicon sequencing method is its compositional information in terms of the relative abundances of the individual members of the community (operational taxonomic units [OTUs], amplicon sequence variants [ASVs], and taxa). However, quantifying different members of complex microbial communities is crucial for differential abundance analysis, such as to better understand the temporal dynamics of microbial communities and, in the case of food microbiology, to identify taxonomic groups that impact quality by causing off-flavors in fermented foods when reaching certain levels [18, 19]. Quantitative real-time PCR (qPCR) is one of the most widely used methods to precisely quantify bacteria in complex ecosystems. A difficulty posed for the quantification of specific taxa in

complex systems is that specific primer systems have to be designed, which can be a very labor-intensive task. In addition, the low throughput of conventional qPCR systems is a limitation that adds to high labor and material costs. The development of high-throughput qPCR (HT-qPCR) has led to a reduction in the work load and material costs (i.e., PCR chemicals) and has opened up new fields of application. These include the investigation of synthetic bacterial soil communities [20], the determination of functional genes in soils [21], the quantification of pathogens in spiked feces and environmental water samples [22], the investigation of microbial diversity in the intestines of piglets [23], and the quantification of bacteriophages of the species *Lactococcus* (*Lc.*) *lactis* and *Leuconostoc* spp. in cheese milk [24].

In food microbiology, qPCR and NGS have been increasingly used in recent years to better understand the microbial composition of various foods [25, 26]. Fermented foods are composed of an often limited number of core species selected by the strictly controlled conditions during the production process and the limited supply of nutrients, pH, and temperature. Therefore, they are particularly suitable for the study of the bacterial communities by qPCR. However, HT-qPCR has only recently been used for the systematic analysis of fermented foods [27], while numerous studies have applied NGS [26, 28–30].

The aim of this study was to compare the performance of a 16S rRNA gene amplicon sequencing approach to a recently developed HT-qPCR method for the analysis of cheese DNA samples. Raclette du Valais protected designation of origin (PDO) cheeses were selected for their higher microbial diversity compared to other cheeses, resulting from the use of raw milk and the low scalding temperature. To this end, we analyzed bacterial community composition in DNA samples from 21 Raclette du Valais PDO cheeses originating from the same number of different cheese producers distributed in the Canton of Valais (Switzerland) using both approaches and compared the results.

Materials and Methods
Sampling
In this study, 21 Raclette du Valais PDO cheese loaves (S01-S21) produced in the same number of different cheese dairies were collected after 120 days of ripening. Twenty of the cheeses were manufactured on the same date, and one cheese (S17) was manufactured 11 days later. Raclette du Valais PDO is a semi-hard, smear-ripened, full fat cheese produced from raw milk and mainly consumed in melted form after a ripening time of at least three months. For the production of Raclette du Valais PDO, the 21 cheese dairies followed the specifications of

the Raclette du Valais AOP association [31]. With regard to the use of starters and additional cultures, a mesophilic starter culture of lactic acid bacteria is added consisting of strains of *Lc. lactis* subsp. *lactis*, *Lc. lactis* subsp. *cremoris,* and *Lc. lactis* subsp. *lactis* biovar *diacetylactis*; and, depending on the production site, the thermophilic lactic acid bacteria *Streptococcus thermophilus* and/or *Lactobacillus helveticus* are occasionally also added to the cheese milk [32].

DNA extraction

Bacterial pellets from cheese were obtained by adding 10 g of cheese to 90 ml modified peptone water (10 g/l peptone from casein, 5 g/l sodium chloride, 20 g/l trisodium citrate dihydrate, pH 7.0) and incubating for 10 min at 40 °C. The sample was then homogenized for 3 min in a Stomacher (Masticator, IUL Instruments, Königswinter, Germany). A 50 μl volume of 10% (w/v) sodium dodecyl sulfate was then added to 10 ml of the homogenate, which was then thoroughly mixed and centrifuged ($4000 \times$ g, room temperature, 30 min). Cell lysis and genomic DNA extraction were performed using the EZ1 DNA Tissue kit and a BioRobot® EZ1 workstation (Qiagen, Hilden, Germany). Briefly, bacterial pellets were resuspended in 250 μl G2 buffer (EZ1 DNA Tissue kit), transferred in 0.5 ml skirted tubes containing 100 mg 0.1 mm low binding zirconium beads (OPS Diagnostics, Lebanon, NJ, USA), and shaken 15 s at medium speed in a bead ruptor (Omni International Inc., Kennesaw, GA, USA). Cell lysates were then processed by the BioRobot® EZ1 workstation. Genomic DNA was eluted in a volume of 100 μl, and the concentration was measured using a NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific, Waltham, MA, USA).

HT-qPCR primers

The primers used for HT-qPCR in this study were described in a previous study [27]. Briefly, 24 target species/subspecies were selected based on a review of the literature and our own preliminary results from the 16S rRNA gene amplicon sequencing of Gruyere and the Raclette du Valais PDO cheeses considered in this study (unpublished data). The selection criteria for the target species were the relative abundance and frequency of detection as well as known impacts on cheese quality.

HT-qPCR standards

The standards for quantification in the HT-qPCR system were produced using standard calibration curves of gBlock™ Gene Fragments (Integrated DNA Technologies, LubioScience, Switzerland), described in detail previously [27]. Copy numbers for quantification were calculated using standard calibration curves ranging from 10^7 to 10^3 copies/μl.

Microfluidic HT-qPCR

HT-qPCR was performed using a 192.24 Dynamic Array integrated fluidic circuit (IFC; Fluidigm Corporation, San Francisco, CA, USA). The assay mix consisted of 3 μl 2× Assay Loading Reagent (Fluidigm Corp.) added to 3 μl primer mix (forward and reverse, 10 μM). A sample pre-mix was prepared by combining 3 μl 2× SsoFast™ EvaGreen® Supermix with low ROX (Biorad, Cressier, Switzerland) and 0.3 μl 192.24 Delta Gene Sample Reagent (Fluidigm Corp.). Finally, 2.7 μl of each sample were added to 3.3 μl sample pre-mix. The IFC was loaded according to the manufacturer's instructions [33]. Briefly, 3 μl of each assay and 3 μl of each sample were distributed to the respective inlet, and the IFC was loaded using the Juno Load Mix 192.24 GE script. The loaded IFC was transferred to the Biomark instrument and run with the GE 192x24 PCR + Melt v2 program, as follows: hot start 95 °C for 1 min, followed by 30 cycles of denaturation at 96 °C for 5 s, and annealing and elongation at 60 °C for 20 s. A melting curve analysis was performed with a temperature increase of 1 °C per 3 s from 60 to 95 °C.

HT-qPCR data analysis

The results from the 192.24 Dynamic Array IFCs were analyzed with the Fluidigm Real-Time PCR Analysis Software version 4.5.2 (Fluidigm Corp.) as described in a previous study [27]. The melting curve peak threshold was set to 0.025 -dRn/dT based on a visual inspection of the baseline fluorescence. All reactions flagged by the Real-Time PCR Analysis Software were interpreted as negative results. The copies/μl of the specific targets were calculated for each reaction using the standard calibration curves, and all reactions below an 800 copies/μl cut-off were interpreted as negative, as recommended by the manufacturer [34]. Average copies/μl were only calculated if at least two of three reactions were positive; otherwise, the results were interpreted as negative.

16S rRNA gene amplicon sequencing

Amplicon libraries were prepared using the unidirectional fusion method (Thermo Fisher Scientific, Waltham, MA, USA). PCR of the V1–V2 16S rRNA gene region was performed in 50 μl reactions using 4 μl of DNA, 0.1 μM primer NGS_ABCxF27 (5′- CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG |Barcode X| AG AGT TTG ATC MTG GCT CAG − 3′) and 0.1 mM primer NGS_trP1_355 (5′- CCT CTC TAT GGG CAG TCG GTG ATG CWG CCT CCC GTA GGA GT − 3′), and 45 μl Platinum™ PCR SuperMix High Fidelity (Thermo Fisher Scientific, Waltham, MA, USA). The

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 4 of 18

amplification was carried out as follows: 94 °C for 2 min, followed by 18 cycles of 94 °C for 30 s, 55 °C for 30 s, and 68 °C for 30 s. All amplicons were purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA) with a bead-to-DNA ratio of 1.8. The quality control and quantification of the amplicon library was performed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and the High Sensitivity DNA Assay. Afterwards, all amplicons were pre-diluted and equimolarly pooled to a 40 pM final library. Template preparation, chip loading, and sequencing were performed according to the manufacturer's instructions using Ion Chef™ System and Ion S5™ System and an Ion530 Chip (Thermo Fisher Scientific, Waltham, MA, USA).

16S rRNA gene amplicon sequencing data analysis

The raw sequences, with an average length of 320 bp, were primer trimmed and quality filtered (maxEE = 15, truncQ = 6, maxN = 0, n = 1e+06, minLen = 100, maxLen = 460) in DADA2 [35]. Amplicon sequence variances (ASVs) were obtained in DADA2 with the parameter POOL = "pseudo." Taxonomic annotation was performed using DAIRYdb v1.2.4 [9] with IDTAXA [36]. Biostatistical analyses were done using the PHYLOSEQ package [37] in R v4.0.2 [38]. Copy number normalization was based on the copy number information available in the Ribosomal RNA Database (rrnDB, version 5.7, January 18, 2021, [39]).

Method comparison data analysis

For the data analysis, we used the following Python packages: Jupyter-notebook v6.2.0 [40] with Python v3.9.2 and IPython v7.21.0 [41], NumPy v1.20.1 [42], seaborn v0.11.1 [43], pandas v1.2.3 [44], SciPy v1.6.1 [45], Matplotlib v3.3.2 [46], statsmodels v0.12.2 [47], and rpy2 v3.4.3. Further, R v4.0.3 [38] and the metacal v0.2.0 package [48] were used for bias estimation (see below). The data analysis was performed as outlined in the htqpcr_ngs_comparison_R.ipynb notebook available in Additional file 1 and the Github repository [49].

Briefly, the taxonomic assignments and number of reads from the 16S rRNA gene amplicon sequencing (NGS) analysis data for the 21 Raclette du Valais PDO samples were extracted from the data set. The most prevalent species in the NGS results were defined as species detected in more than 30% of the cheese samples. Relative species abundance was calculated for each cheese DNA sample considering all members of the community to create plots representing the community composition. HT-qPCR analysis was performed using the HTqPCR_dataparser.py script. Further, the data for the two *Lc. lactis* subspecies from HT-qPCR were grouped to *Lc. lactis*. For the bias estimation, the data from both methods were filtered to consider only the 15 bacterial species/groups measured by both methods (shared positive). The NGS data were defined as the observed category, and only read counts for species also detected by the HT-qPCR approach (reference) were included. Further, pseudocounts (= 1) were added in the observed data (NGS) for species only detected in the reference data (HT-qPCR). Bias estimates were calculated using metacal with the number of reads and copies of the 15 investigated bacterial species/groups measured by both methods as input. A corrected data set was made by grouping the data for *Lactiplantibacillus pentosus* and *Lactiplantibacillus plantarum* for NGS and for *L. plantarum* and *Lactiplantibacillus paraplantarum* for the HT-qPCR to a common *L. plantarum* group category.

Bias estimation

The model for bias estimation is described in detail in [48]. Briefly, the assumption of this bias model is that the bias is caused by the different efficiencies for the given measurement (relative or absolute abundance) of different species. The bias estimates are calculated from taxon proportions to make the bias independent of the sample's composition. The systematic difference between measurements from different methods can be estimated by the difference in their biases. If the actual composition is not known, but a reference composition is considered as the true composition, this differential bias is equivalent to the bias of the method under investigation. A point estimate of the bias (the ratio of the efficiency of a species to the geometric mean efficiency of all species) for each species with known (reference) abundance can then be calculated for the samples. Geometric standard errors were estimated from 1000 bootstrap replicates.

Reference sequence alignments

Representative genomes of the reference species were downloaded from the National Center for Biotechnology Information (NCBI). The 16S rRNA gene sequences were extracted using Barrnap V 0.9 [50]. Non-redundant sequences of the V1–V2 region of the 16S rRNA gene were aligned using PRANK V .150803 [51].

Construction of phylogenetic tree from reference sequences and ASVs

The reference 16S rRNA gene sequences for the *L. plantarum* group species were extracted from the DAIRYdb v1.2.4 fasta file [52]. ASVs assigned to the *L. plantarum* group species were filtered. The sequences of the V1–V2 region of the 16S rRNA gene were aligned using PRANK V .150803, and the resulting multiple sequence alignment was subjected to a rapid bootstrap analysis using RAxML

Dreier *et al. BMC Microbiology*　　　(2022) 22:48

Page 5 of 18

[53]. The best-scoring maximum likelihood tree was visualized using iTOL [54].

Results and Discussion

HT-qPCR and 16S rRNA gene amplicon sequencing results

The average sequencing depth for 16S rRNA gene amplicon sequencing was 471,184 reads (range: 361496–632,269). In total, 9,894,860 reads were classified to 233 ASVs. These ASVs were assigned to 47 different sequence groups, and 45 of these were classified to the species level, while for two ASV groups only a classification to the family level (*Ruminococcaceae, Streptococcaceae*) was possible. Four core species were detected in all 21 cheese samples. These species were *Lacticaseibacillus paracasei*, *Lc. lactis*, *L. helveticus*, and *S. thermophilus*. Ten species were present in more than 80% of the samples and belonged to either the *Lactobacillaceae* or the *Streptococcaceae* families. The 21 most prevalent species (occurring in at least 30% of cheeses) represented on average 99.96% (range 99.79–100%) of all the reads and included species

from the *Lactobacillaceae, Leuconostocaceae, Enterococcaceae,* and *Streptococcaceae* families (Fig. 1).

The HT-qPCR system consisted of specific primer pairs targeting 24 different bacterial species/subspecies, of which a total of 17 species/subspecies were quantified in at least one of the cheese DNA samples (Fig. 2). *Lc. lactis* subsp. *lactis*, *Lc. lactis* subsp. *cremoris,* and *L. paracasei* were detected in all samples, while *S. thermophilus* and *L. plantarum* were detected in all but one sample (sample S10). *Lc. lactis* subsp. *lactis* was the dominant subspecies in all samples. Surprisingly, *L. helveticus* was detected only in five samples by HT-qPCR. Three additional species (*Lentilactobacillus parabuchneri, Loigolactobacillus coryniformis,* and *Latilactobacillus curvatus*) were detected in more than 80% (17) of the samples, while *L. paraplantarum* was detected in 62% (13) of the cheese samples. The seven species in the HT-qPCR system that were not detected in any of the cheeses examined corresponded to *Clostridium tyrobutyricum, Enterococcus durans, Enterococcus faecium, Lacticaseibacillus casei,*
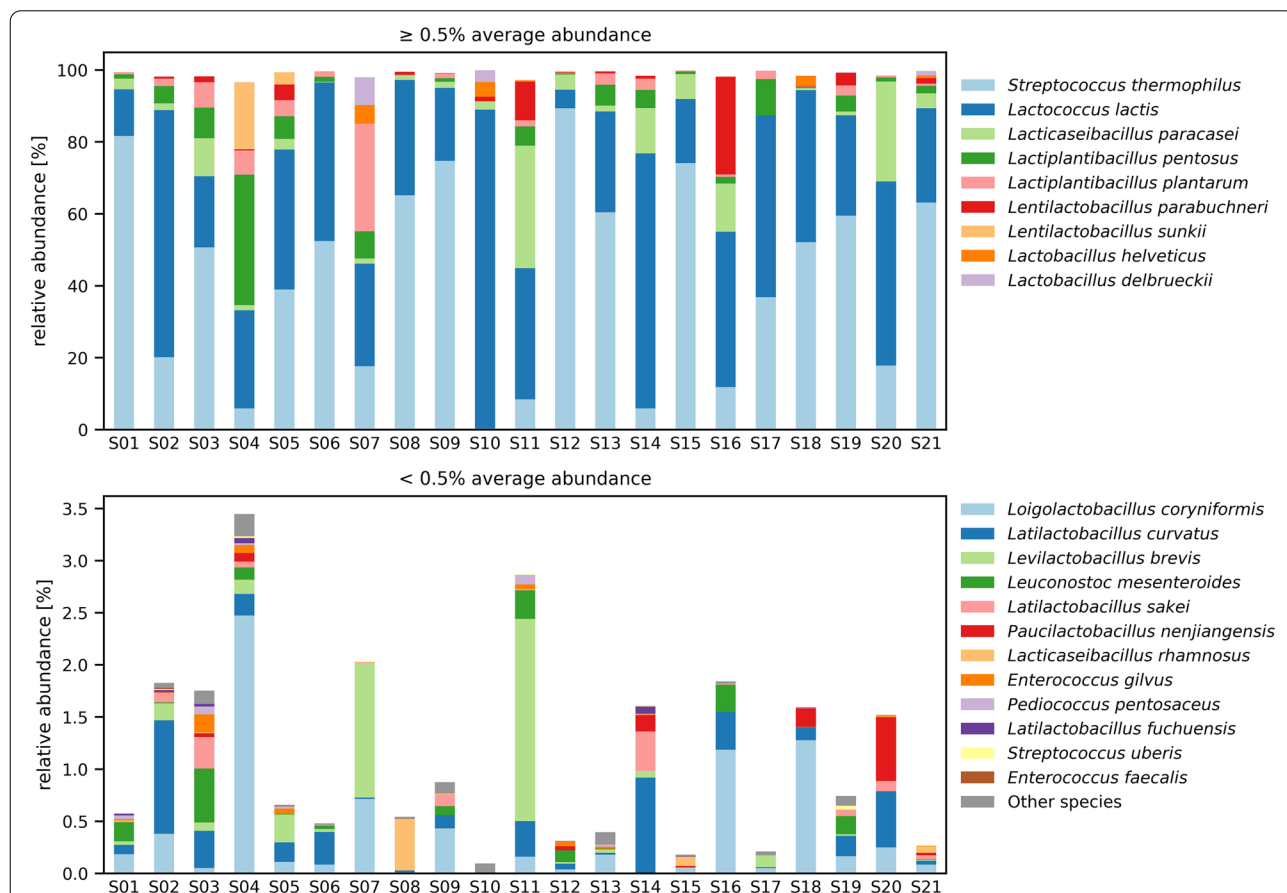
Fig. 1 Bacterial community composition determined by 16S rRNA gene amplicon sequencing. Species detected in Raclette du Valais PDO cheese DNA samples (S01-S21) with an average relative abundance above 0.5% are shown in the upper panel and the species with a lower average abundance in the lower panel. The relative abundance of the 21 species detected in more than 30% of the samples are depicted with the species name; the other 26 taxa were classified as other species

log copies/µl 3 4 5 6 7 8

Species	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21
L. lactis subsp. *lactis*	5.82±0.25	6.17±0.04	5.63±0.02	5.51±0.02	5.97±0.04	6.21±0.08	5.79±0.07	6.14±0.16	5.66±0.29	6.56±0.1	5.67±0.04	5.64±0.02	6.12±0.03	5.65±0.07	5.79±0.05	5.6±0.01	6.26±0.22	6.47±0.14	5.81±0.02	5.59±0.04	5.96±0.07
S. thermophilus	6.6±0.22	5.8±0.03	6.2±0.03	5.03±0.03	6.03±0.02	6.37±0.05	5.66±0.05	6.59±0.08	6.31±0.25		5.14±0.04	6.93±0.02	6.42±0.02	4.83±0.03	6.54±0.02	5.23±0.03	6.28±0.19	6.52±0.14	6.17±0.02	5.23±0.04	6.3±0.05
L. paracasei	5.42±0.21	4.9±0.03	5.73±0.04	4.55±0.03	5.09±0.04	4.18±0.13	4.72±0.06	5.11±0.07	4.85±0.33	5.18±0.1	5.83±0.05	5.85±0.01	5.11±0.02	5.22±0.03	5.63±0.03	5.29±0.04	2.96±0.03	4.83±0.15	4.71±0.03	5.55±0.04	5.4±0.05
L. plantarum	5.12±0.1	5.46±0.02	5.82±0.03	5.92±0.02	5.65±0.02	5.32±0.05	6.08±0.05	4.36±0.02	5.1±0.13		5.0±0.04	4.98±0.02	5.85±0.01	5.09±0.05	4.9±0.04	4.61±0.05	6.11±0.11	4.86±0.06	5.5±0.04	4.43±0.1	5.27±0.06
L. lactis subsp. *cremoris*	4.49±0.86	5.24±0.15	4.89±0.02	5.04±0.02	4.74±0.03	5.63±0.04	4.45±0.07	5.75±0.2	3.95±1.03	4.31±0.27	4.69±0.01	5.01±0.01	5.02±0.03	4.92±0.08	3.98±0.08	4.17±0.15	5.26±0.79	5.35±0.45	4.28±0.05	4.59±0.03	5.04±0.1
L. parabuchneri	3.23±0.18	4.49±0.07	5.06±0.01	3.95±0.08	5.37±0.03		3.0±0.01 (2/3)	5.0±0.03	4.02±0.35	5.28±0.1	5.47±0.05	4.6±0.02	4.78±0.01	4.48±0.09	3.91±0.05	5.73±0.01		4.59±0.13	5.22±0.02		5.1±0.04
L. coryniformis	3.87±0.24	4.21±0.04	3.43±0.19	4.74±0.02	3.73±0.13	3.74±0.21	4.37±0.15	3.02±0.01 (2/3)	4.24±0.36		3.51±0.16	3.82±0.07	4.17±0.09		3.44±0.12	4.28±0.09	3.74±0.26	5.26±0.1	3.82±0.12	3.43±0.13	3.67±0.11
L. curvatus	4.18±0.24	4.98±0.03	4.49±0.06	3.8±0.1	4.31±0.04	4.61±0.06	3.49±0.27 (2/3)	3.57±0.07	4.01±0.29		4.13±0.05	4.36±0.04	3.49±0.25	4.32±0.03		4.12±0.05		4.49±0.23	4.15±0.08	4.16±0.08	3.52±0.14
P. pentosaceus	4.25±0.64		4.65±0.02	4.18±0.02	4.09±0.11	3.31±0.09 (2/3)			3.76±0.68		4.28±0.04	3.14±0.18 (2/3)	4.35±0.04			3.18±0.07	3.27±0.5	3.76±0.32	3.82±0.18		
L. paraplantarum	4.73±0.18	4.58±0.06	3.96±0.05	3.97±0.04	3.65±0.08	4.09±0.1			4.22±0.35					3.22±0.19		3.86±0.13	3.27±0.21	3.28±0.06	3.17±0.18		3.39±0.18
L. sakei	3.26±0.25	3.8±0.14	4.36±0.07	3.31±0.18		3.17±0.24 (2/3)			3.81±0.34					3.84±0.17					3.66±0.28	3.18±0.1	3.4±0.16
L. mesenteroides	4.2±0.43		4.47±0.06	3.58±0.11		3.59±0.16 (2/3)			3.46±0.67		3.83±0.08	4.32±0.03				3.68±0.13			3.84±0.11		
L. brevis	3.05±0.05 (2/3)	3.5±0.29		3.5±0.08	3.93±0.05		4.46±0.07				4.45±0.07	3.05±0.15 (2/3)									
L. helveticus							5.15±0.04		5.33±0.07	3.67±0.09								5.55±0.09			4.52±0.08
L. rhamnosus			3.05±0.06 (2/3)				4.73±0.04	3.1±0.12 (2/3)						3.72±0.12							3.68±0.11
L. delbrueckii							4.97±0.07		4.93±0.11									3.71±0.2			4.42±0.08
E. faecalis		3.49±0.33 (2/3)																			

Fig. 2 Heatmap of HT-qPCR results. The heatmap annotation depicts the average log copies/µl and the standard deviation of technical triplicates. When not all samples were positive, the number of positive samples out of the total number of samples is given in brackets

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 7 of 18

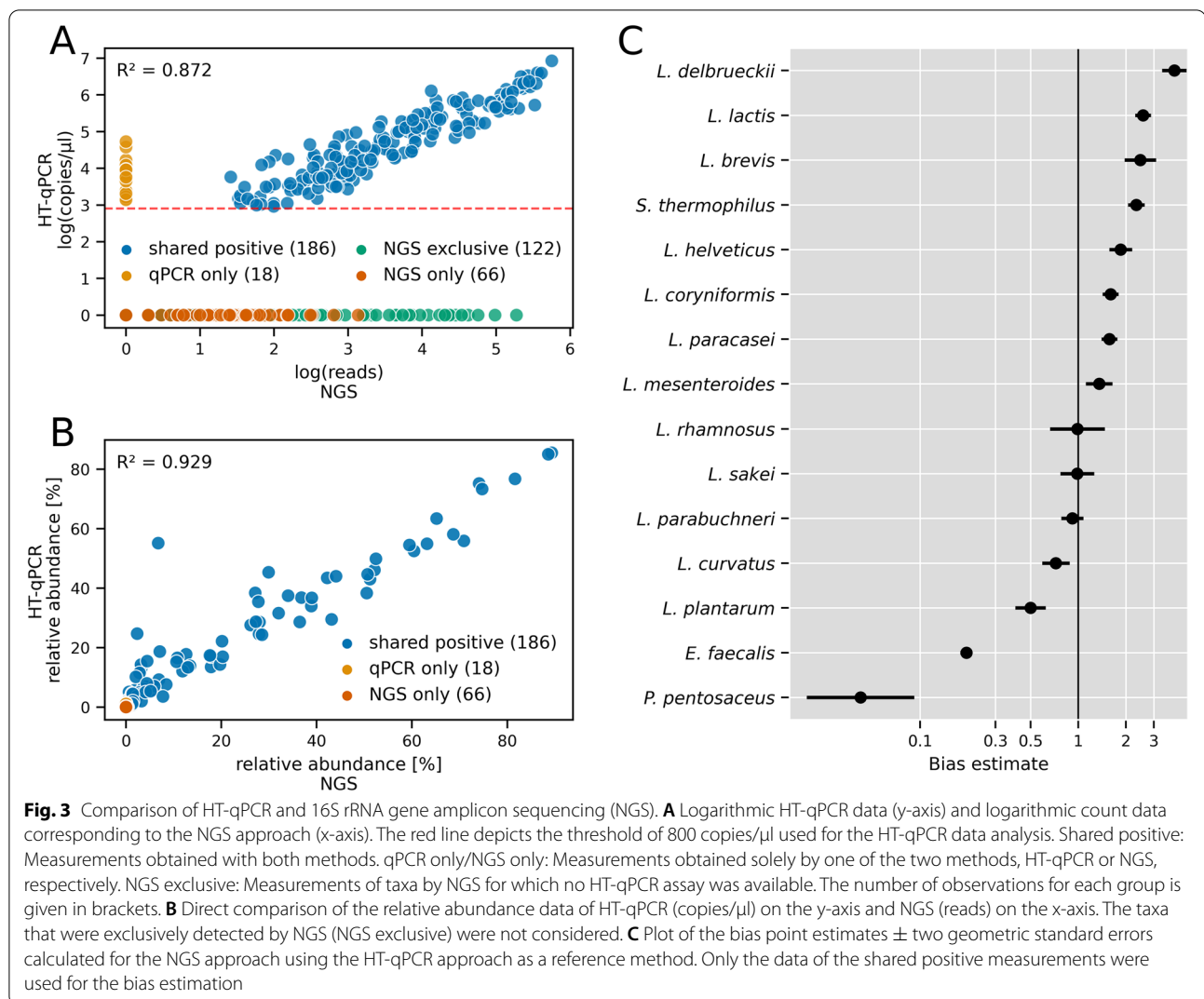*Limosilactobacillus fermentum*, *Pediococcus acidilactici*, and *Propionibacterium freudenreichii*.

The sampled Raclette du Valais PDO originated from commercial batches of good-quality cheeses and thus did not exhibit any sensory-perceptible quality defects. Therefore, the absence or at least very low relative abundance of bacterial species responsible for cheese quality defects, such as *C. tyrobutyricum* (butyric acid fermentation) or *E. durans* and *E. faecium* (potential tyramine producers), was expected. Moreover, in the production of Raclette du Valais PDO, no adjunct cultures containing *L. casei*, *L. fermentum*, *P. acidilactici,* or *Propionibacterium freudenreichii* are used, therefore the detection of these species depends on whether they were present in the raw milk.

To our knowledge, no study has investigated Raclette du Valais PDO cheese using 16S rRNA amplicon sequencing and HT-qPCR yet. Other semi-hard cheese varieties made from raw milk have been investigated by 16S rRNA gene amplicon sequencing, but only few studies report the community compositions at a species level resolution. In two Raclette-type cheeses made from raw milk, analyzed with the same NGS approach as in the present study, the dominant species were *Lc. lactis*, *L. plantarum/pentosus* and *Weissella paramesenteroides*, with a relative abundance of *Lc. lactis* exceeding 50% [29]. In the present study *W. paramesenteroides* was also detected in five samples with a low abundance (<0.085%). A recent study identified *Lactobacillus delbrueckii*, *Lacticaseibacillus rhamnosus*, *L. casei*, *L. helveticus* and *L. fermentum* as the most abundant and prevalent species in Grana Padano cheeses [55]. In contrast to Raclette du Valais the scalding temperature is higher (56 °C) and the natural whey starters are dominated by *S. thermophilus*, *L. delbrueckii* and *L. helveticus* and lower proportions of *L. fermentum* [56]. Ten species, namely, *S. thermophilus*, *Lc. lactis*, *L. rhamnosus*, *Latilactobacillus sakei*, *L. coryniformis*, *Pediococcus pentosaceus*, *W. paramesenteroides*, *L. plantarum*, *L. (para-)casei* and *Weissella hellenica* (sample S20 only) with an average relative abundance above 0.5% in three samples of Danish cheeses made from raw milk ripened for 56 days [57], were also found in our study of Raclette du Valais. In contrast to Grana Padano, the scalding temperature used for the manufacture of this cheese was 39 °C, which is more comparable to the temperature used for Raclette du Valais manufacture (36 °C). *Ligilactobacillus acidipiscis* and *Staphylococcus saprophyticus* were taxa of the indigenous microbiota exclusively found in the Danish cheeses. These results support the observations of many studies showing that parameters such as the type and origin of milk, milk treatment, and the type of ripening significantly influence the microbiota in ripened cheese (reviewed in [58, 59]).

Comparison of HT-qPCR and 16S rRNA gene amplicon sequencing results

The HT-qPCR analysis results represent, after comparison with a standard curve, the number of copies of the species-specific single copy gene per µl of sample, while the amplicon sequencing results correspond to the relative abundance of taxa based on the number of reads with respect to the total number of reads of the corresponding V1–V2 16S rRNA gene region. The fundamental differences in the resulting data (absolute or relative abundances) and the data analysis (standard or compositional) make a direct comparison between the two methods challenging. The first attempt to compare the performance of the two methods qualitatively was based on a comparison of the measured copy numbers versus the number of NGS reads (Fig. 3A). Given that the HT-qPCR system contained specific primer pairs that were able to discriminate between the subspecies of *Lc. lactis* (subsp. *lactis* and subsp. *cremoris*), while the amplicon sequencing was not able to discriminate these subspecies, the data for the two subspecies were pooled to account for the total number of *Lc. lactis* for the comparison of the methods. The data points in Fig. 3A were divided into four groups. The group "shared positive" represents measurements for the 15 species that were covered and detected by both methods. The second and third groups ("qPCR only" and "NGS only") included measurements in which the same 15 species were detected either by HT-qPCR or NGS. The measurements for all other taxa that were not covered by the selected HT-qPCR assays were classified as "NGS exclusive." The detection of a larger number of exclusive taxa using the 16S rRNA gene amplicon sequencing method was expected given the non-targeted nature of the NGS approach. However, in the case of the two species *L. paraplantarum* and *P. pentosaceus* that were detected solely by HT-qPCR, the unexpected outcome may indicate errors or bias in the analysis of the NGS results. For the log-transformed data (Fig. 3A) and the relative abundance data (Fig. 3B) of the "shared-positive" group, positive linear correlations ($R^2 = 0.872$ and $R^2 = 0.929$, respectively) were observed. The relative abundance data in Fig. 3B indicate that the qualitative disagreement between the methods was mainly due to species with low relative abundance, which were detected only by the NGS method.

A summary of the frequencies of the detection of the most prevalent species by both methods is shown in Table 1. In 66 cases, NGS exclusively detected species also targeted by HT-qPCR ("NGS only"), and the relative abundances for these species were in a range between 0.0002 and 0.322%. For example, *L. helveticus* was detected in all samples by NGS, while HT-qPCR did not detect *L. helveticus* in 16 samples with low relative

Dreier *et al. BMC Microbiology* (2022) 22:48

Page 8 of 18

Fig. 3 Comparison of HT-qPCR and 16S rRNA gene amplicon sequencing (NGS). **A** Logarithmic HT-qPCR data (y-axis) and logarithmic count data corresponding to the NGS approach (x-axis). The red line depicts the threshold of 800 copies/µl used for the HT-qPCR data analysis. Shared positive: Measurements obtained with both methods. qPCR only/NGS only: Measurements obtained solely by one of the two methods, HT-qPCR or NGS, respectively. NGS exclusive: Measurements of taxa by NGS for which no HT-qPCR assay was available. The number of observations for each group is given in brackets. **B** Direct comparison of the relative abundance data of HT-qPCR (copies/µl) on the y-axis and NGS (reads) on the x-axis. The taxa that were exclusively detected by NGS (NGS exclusive) were not considered. **C** Plot of the bias point estimates ± two geometric standard errors calculated for the NGS approach using the HT-qPCR approach as a reference method. Only the data of the shared positive measurements were used for the bias estimation

abundances (0.004–0.043%). Similarly, *L. delbrueckii* was exclusively detected by NGS in 15 cheese DNA samples with relative abundances between 0.002–0.025%. In contrast, *L. paraplantarum* and *P. pentosaceus* were detected in a higher number of samples by HT-qPCR than by NGS. *L. paraplantarum* was detected in 13 samples by HT-qPCR with abundances in a range between 1472 and 53,306 copies/µl, whereas this species was never detected by NGS in any of the analyzed samples. In five samples, *P. pentosaceus* was exclusively detected by HT-qPCR with a range of 1389–6542 copies/µl. Reciprocally, *P. pentosaceus* was exclusively detected by NGS in one sample (relative abundance 0.0055%).

The relatively high detection limit of 800 copies/µl is most likely responsible for the inability of the HT-qPCR method to detect low abundance species. The number of target gene copies presumably also has an influence on sensitivity. The HT-qPCR system targets specific

single-copy genes, while according to the rrnDB database, most of the prevalent species contain about five copies (range: 3–9) of the 16S rRNA gene (Additional file 2). The sensitivity of the HT-qPCR assays is further limited by the nanoliter-scale reactions used in the microfluidic qPCR system compared to standard qPCR methods. Apart from these differences for low abundance species, the 15 species included in the method comparison represented 93.84% (range: 44.69–99.93%) of the reads from the NGS analysis (Table 2), indicating that the most dominant members of the microbial population in the 21 Raclette du Valais PDO samples was covered by both approaches. Samples for which the coverage was below the average (S03, S04, S05, S07, and S17) all showed above-average relative abundances of *L. pentosus*. In addition, sample S04 also showed an above-average relative abundance of *Lentilactobacillus sunkii*. Besides these two species, the other species not targeted

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 9 of 18

Table 1 Detected species and average relative abundance for the HT-qPCR and 16S rRNA gene amplicon sequencing (NGS) in 21 Raclette du Valais PDO cheese samples

	qPCR			NGS		
	Count	Avg. abund. [%]	SD	Count	Avg. abund. [%]	SD
Lacticaseibacillus paracasei	21	8.26	10.59	21	6.34	9.10
Lactococcus lactis	21	32.79	18.17	21	37.18	20.40
Streptococcus thermophilus	20	42.03	25.51	21	42.24	28.20
Lactiplantibacillus plantarum	20	12.00	14.76	20	3.47	6.56
Loigolactobacillus coryniformis	19	0.60	0.95	19	0.42	0.62
Lentilactobacillus parabuchneri	18	4.96	9.28	20	2.68	6.27
Latilactobacillus curvatus	18	0.77	0.85	19	0.26	0.30
Lactobacillus helveticus	5	3.28	2.48	21	0.63	1.47
Leuconostoc mesenteroides	9	0.31	0.22	14	0.13	0.15
Lactobacillus delbrueckii	4	1.58	1.54	19	0.66	1.89
Levilactobacillus brevis	7	0.47	0.61	16	0.26	0.55
Latilactobacillus sakei	10	0.24	0.25	12	0.10	0.12
Pediococcus pentosaceus	13	0.41	0.43	9	0.03	0.03
Lactiplantibacillus pentosus	0			20	5.21	7.91
Lacticaseibacillus rhamnosus	5	0.24	0.36	13	0.05	0.14
Enterococcus gilvus	0			15	0.03	0.05
Lactiplantibacillus paraplantarum	13	0.38	0.42	0		
Lentilactobacillus sunkii	0			10	2.23	5.85
Streptococcus uberis	0			10	0.01	0.01
Latilactobacillus fuchuensis	0			9	0.02	0.02
Enterococcus faecalis	1	0.11		7	0.00	0.00
Paucilactobacillus nenjiangensis	0			8	0.14	0.20
Other species	0			20	0.05	0.06

The relative abundance of the 21 species detected in more than 30% of the samples by NGS and *Lactiplantibacillus paraplantarum* exclusively detected by HT-qPCR are depicted with the species name; the other 26 taxa were classified as other species

by the HT-qPCR system accounted for only 0.13% (range: 0.01–0.63%) of the NGS reads. Neither *L. pentosus* nor *L. sunkii* were included in the HT-qPCR system since no validated primers for these species were available. For future studies, it would be beneficial to design specific primers for these species to enable the quantification of these common species in cheese using the HT-qPCR approach. *L. pentosus* has already been isolated from milk and cheese [58, 60]. *L. sunkii* was originally isolated from sunki, an unsalted Japanese fermented food, and has already been detected in kefir biofilms [61, 62]. However, to our knowledge, the first detection of *L. sunkii* in cheese was reported only recently in an NGS study of Grana Padano cheese [55].

Bias estimation

Over the 21 investigated cheese samples, the two methods show a high degree of qualitative agreement (detected or undetected) and a strong correlation between the measurements for the species covered by both approaches (Fig. 3 A and B). However, the correlation of relative abundance data for the same samples measured by two different methods is not a suitable indicator for the agreement or disagreement of the methods, since a high correlation can be expected for two methods measuring similar properties in the same samples. To examine the differences between the two methods in more detail, point estimates of the bias for the 15 shared species were calculated for the NGS approach using the HT-qPCR approach as a reference method (Fig. 3C). A bias estimate value above 1 indicates an increased efficiency for NGS, while values below 1 indicate a decreased efficiency for the measurement of the species compared to the reference method (HT-qPCR). A strong positive bias was observed for *L. delbrueckii, Lc. lactis, Levilactobacillus brevis*, and *S. thermophilus*, while a strong negative bias was observed for *L. plantarum, Enterococcus faecalis,* and *P. pentosaceus.*

Identification of possible causes for negative bias

Potential biases can be introduced at every step of the NGS protocols, from nucleic acid extraction, library

Table 2 Relative abundance data of species detected by 16S rRNA gene amplicon sequencing

	Shared species		Other species		L. pentosus	L. sunkii
	[%]	[n]	[%]	[n]	[%]	[%]
S01	98.73	13	0.03	3	1.24	0.01
S02	95.06	12	0.09	6	4.83	0.02
S03	91.16	14	0.36	6	8.47	
S04	44.69	14	0.44	11	36.27	18.6
S05	90.24	13	0.07	4	6.33	3.36
S06	98.62	12	0.01	5	1.37	
S07	92.47	12	0.01	1	7.52	
S08	99.93	10	0.02	3	0.05	
S09	98.89	12	0.11	3	1.00	
S10	99.91	7	0.09	1		
S11	94.59	15	0.05	3	5.36	
S12	99.31	13	0.09	5	0.33	0.27
S13	94.06	12	0.14	4	5.81	
S14	94.76	13	0.23	5	5.01	
S15	99.26	10	0.03	5	0.71	
S16	98.11	11	0.02	4	1.86	
S17	89.88	11	0.04	2	10.08	
S18	99.5	11	0.18	3	0.31	
S19	95.44	14	0.13	6	4.43	
S20	98.24	10	0.63	5	1.12	0.01
S21	97.85	13	0.03	7	2.11	

Percentage of reads assigned to species also covered by HT-qPCR (shared) and other species that were not covered by HT-qPCR. The two species, *Lactiplantibacillus pentosus* and *Lentilactobacillus sunkii,* with the largest overall abundance are shown separately

preparation, and sequencing, to the bioinformatics analysis [1, 48]. McLaren et al. [48] have shown, using mock communities analyzed by NGS, that the largest influence on the total bias originates from DNA extraction, followed by PCR. In our study, we used the same DNA samples for the measurements by both methods, therefore bias due to DNA extraction did not contribute to the observed total bias. We performed a bioinformatics analysis to identify possible causes of the decreased efficiency of NGS for *L. plantarum* and *P. pentosaceus.* The analyses were not repeated for *E. faecalis* as the bias estimate was based on just a single measurement. Alignments of the V1–V2 region of the 16S rRNA gene sequences from the representative genomes have shown one or two nucleotides difference between *L. plantarum* and *L. pentosus* and only three or four nucleotides difference for *L. paraplantarum* (Supplementary Fig. S1 A, Additional file 3). Multiple copies of the 16S rRNA gene in the representative genomes of *L. plantarum* and *L. paraplantarum* were not identical and contained single nucleotide polymorphisms (SNPs). The DAIRYdb (v.1.2.4), used by IDTAXA for taxa assignment, contained four 16S rRNA gene reference sequences spanning the entire length of the V1–V2 region, two identical sequences for

L. pentosus and one each for *L. plantarum* and *L. paraplantarum* (Supplementary Fig. S1 B, Additional file 3). The references for *L. plantarum* and *L. pentosus* contained only a single nucleotide transition. In the phylogenetic tree (Supplementary Fig. S1 C, Additional file 3) of the ASVs, two distinct clades could be identified, one for *L. paraplantarum* (purple) and one for *L. plantarum* and *L. pentosus* (green), including the genomic reference sequences and type strain sequences from the DAIRYdb. However, considering the high similarity of the reference sequences and the intra-strain SNPs in the representative genomes, it is likely that the assignment to *L. pentosus* or *L. plantarum* was based on single nucleotide differences. Nevertheless, it remains unclear why *L. paraplantarum* was never assigned by DAIRYdb-IDTAXA in the analysis pipeline.

These findings indicate the inability to differentiate *L. pentosus*, *L. plantarum,* and *L. paraplantarum* based on the selected primers for the V1–V2 variable region of the 16S rRNA gene as a source of underestimation bias using 16S rRNA gene sequencing. The difficulty in differentiating the species of the *L. plantarum* group has already been identified earlier [63–65]. Torriani et al. [66] reported that the partial sequences of the *recA, dnaK, tuf,*

Dreier *et al. BMC Microbiology*　　　(2022) 22:48

Page 11 of 18

*hsp*60, and *pheS* genes allow a better differentiation of *L. plantarum, L. pentosus,* and *L. paraplantarum.*

Further investigations were undertaken to clarify whether other regions of the 16S rRNA gene would be better suited for the differentiation of these species. Primers targeting the V3–V4 variable regions of the 16S rRNA gene, which were used in recent studies of microbial populations in milk and cheese [30, 67], are even more problematic as this region displays 100% nucleotide identity for *L. plantarum, L. pentosus,* and *L. paraplantarum* in the representative genomes (Supplementary Fig. S2, Additional file 3). To prevent biased microbiota data, studies in fermented foods that rely solely on species identification based on 16S rRNA gene sequences should mention this limitation if *L. plantarum* group species are included or, alternatively, try to differentiate these species by additional analysis, such as multiplex PCR.

Regarding the investigation of the strong bias for *P. pentosaceus*, the potential reasons for PCR bias were examined. An alignment of the primer regions showed that only the sequence of *P. pentosaceus* had a potential mismatch at position 12 of the NGS_ABCxF27 primer (Supplementary Fig. S3, Additional file 3). The wobble base (M) at position 12 of the primer represents an adenine or cytosine, while the *P. pentosaceus* sequence at this position contains a thymine. Since it is a single mismatch and is not located at the 3′ end, it certainly does not prevent amplification but most likely can reduce primer efficiency. Notably, in one sample (S14), *P. pentosaceus* was detected by NGS but not by HT-qPCR. The relative abundance of *P. pentosaceus* in this sample was very low (0.0055%), and the total copies/µl of cheese sample S14 was low compared to the other samples. Looking at the qPCR raw data, we observed weak fluorescent signals in some reactions (technical triplicates) containing the *P. pentosaceus* assay;
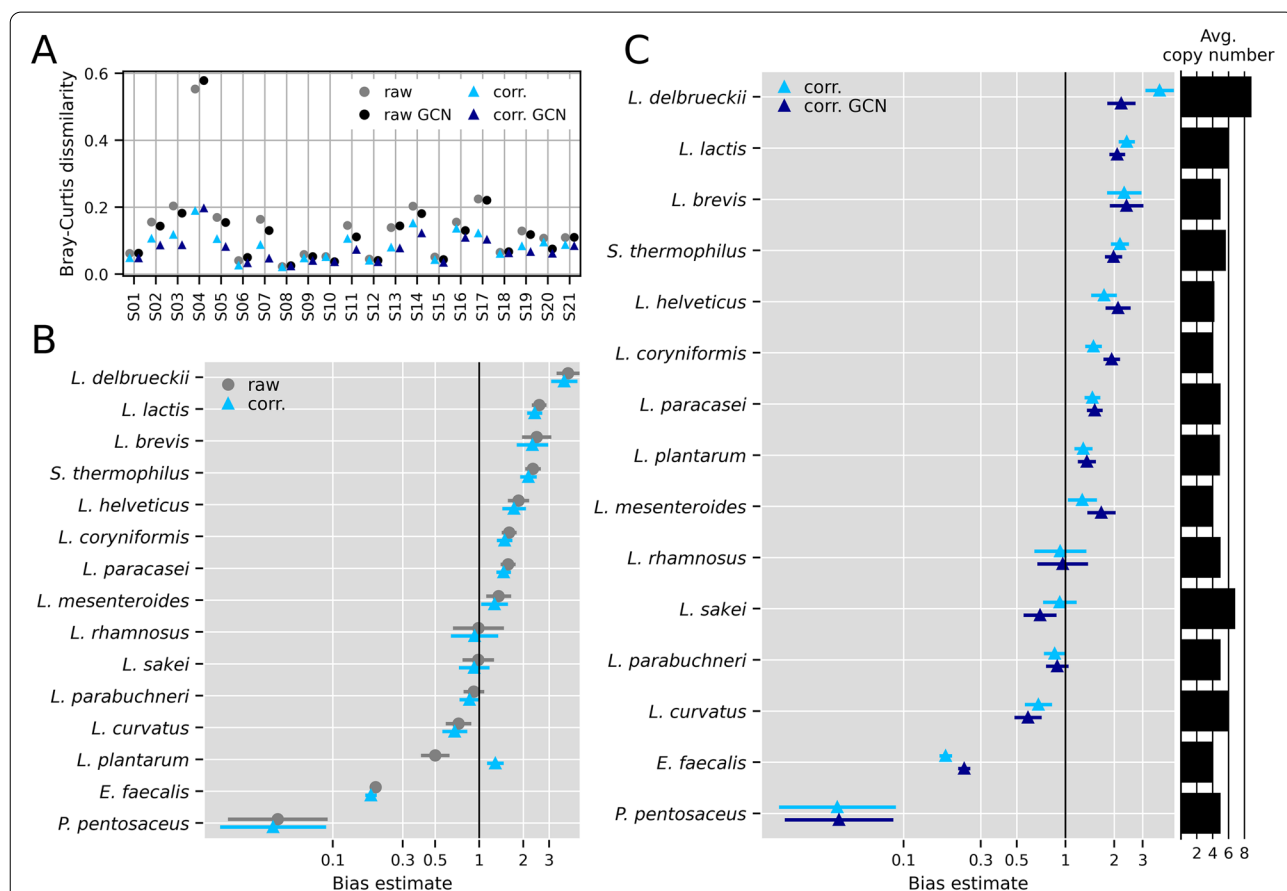
Fig. 4 Bias estimates for raw and corrected data sets. **A** Bray-Curtis dissimilarity of relative abundance data of 16S rRNA gene amplicon sequencing (NGS) to HT-qPCR relative abundance data (reference). Depicted are the dissimilarities for the raw and the corrected *Lactiplantibacillus plantarum* group data sets without and with gene copy number normalization (GCN) for the 16S rRNA gene. **B** Bias point estimates ± two geometric standard errors before and after the correction of the assignments for the *L. plantarum* group species. **C** Bias point estimates ± two geometric standard errors after the correction of the assignments for the *L. plantarum* group species, with (corr. GCN) and without (corr.) 16S rRNA gene copy number correction

however, it was below the 800 copies/µl cut-off value used to improve the signal-to-background noise ratio.

Possible factors for positive bias were not investigated in detail here, but presumably the copy number of the 16S rRNA gene and PCR bias have an influence on the positive total bias. In addition, it must also be considered that even though we have chosen HT-qPCR as a reference for this comparison, the method is not independent of its own inherent bias.

Bias correction

Since the 16S rRNA gene sequencing analysis work-flow was not able to differentiate between the three species of the *L. plantarum* group, those were combined into an *L. plantarum* group as a post-analysis bioinformatics correction. After this correction, the 15 investigated species (now including *L. pentosus*) accounted for 98.81% (range: 80.96–99.99%) of the reads of the NGS analysis. Only samples S04 and S05 had a coverage below the average, due to a high relative abundance (18.6 and 3.4%, respectively) of *L. sunkii* (Additional file 4). Another approach that has been investigated for partial correction of PCR bias is gene copy number normalization (GCN) of the 16S rRNA gene.

After the *L. plantarum* group correction, the Bray-Curtis dissimilarity relative to the HT-qPCR approach was decreased for all samples except S08 and S10 (Fig. 4A). Samples S08 and S10 were the only samples with a relative abundance of *L. pentosus* below 0.05%. Further, the Bray-Curtis dissimilarity was lower for the 16S copy number correction (black circles) compared to the raw data (gray circles) in most samples; only in samples S04, S06, S08, and S18 was the Bray-Curtis dissimilarity higher. A lower Bray-Curtis dissimilarity was calculated for the data with combined corrections for the *L. plantarum* group species and GCN (dark blue triangles) compared to the raw data. The bias for *L. plantarum* was noticeably improved for the *L. plantarum* group correction, while the bias for the other species barely changed (Fig. 4B). The bias changed from a decreased efficiency for the measurement of *L. plantarum* to an increased efficiency for the *L. plantarum* group species. This may be partially explained by the proportion of *L. pentosus* since *L. pentosus* was not measured by HT-qPCR.

For the GCN, the number of reads measured by NGS was divided by the average number of copies for each species. This approach had only a minor effect on the estimated biases (Fig. 4C). The average 16S rRNA copy number for the 15 investigated species was 5.3 copies; therefore, for species with higher copy numbers, the efficiency of the measurement decreased (*S. thermophilus, Lc. lactis, L. curvatus, L. sakei, L. delbrueckii*), while for species with lower copy numbers, the efficiency

increased. The 16S rRNA GCN had only a major influence on the bias for species with a high average number of copies, namely *L. delbrueckii* (avg. copies: 8.9). McLaren et al. [48] have previously reported that the total bias was poorly explained by copy number correction for the mock communities used in their study. Improving the predictions for the composition of microbial communities based on 16S rRNA GCN, apart from mock communities, is still an unsolved problem [68]. Difficulties include, for example, that predictions of 16S rRNA copy numbers can be inaccurate and strongly differ between prediction tools for taxa with unknown numbers of copies of the 16S rRNA gene [69]. Other unresolved issues include varying copy numbers within the same genus or the intra-genomic heterogeneity of the 16S rRNA gene [70, 71].

Potential and limitations

The strength of the HT-qPCR approach lies in the fast and reliable analysis of samples with a known composition. The strength of NGS for exploratory purposes is very evident, as for example, bacterial species previously not associated to the cheese microbiota where discovered when NGS was applied to artisanal cheese samples [72]. Moreover here we show that overall, the results obtained by NGS and HT-qPCR mostly agreed for the relative abundance of a set of 15 shared bacterial species in 21 cheese DNA samples after the bioinformatics corrections for the *L. plantarum* group species. Unweighted pair group method with arithmetic mean (UPGMA) linkage based on Bray-Curtis dissimilarity clustered the measurements of NGS and HT-qPCR together for most of the 21 samples (Fig. 5). Only for cheese DNA samples S13, S19, and S21 did the results between the two methods diverge.

The present study aimed to evaluate the accuracy of a 16S rRNA gene amplicon sequencing approach in cheese by comparing it to absolute abundance data of selected taxa. During the early days of the use of 16S rRNA gene amplicon sequencing for the investigation of the cheese microbiota, the method was often compared to culture-dependent methods or qualitative culture-independent methods such as denaturing gradient gel electrophoresis [73–78]. However, advances in NGS technologies regarding optimization of the most discriminative 16S rRNA gene regions, primers, longer reads, and curated databases for specific ecosystems has increased the taxonomic resolution to the species level. Comparisons with quantitative culture-dependent methods are often limited in terms of species level resolution and bacteria in a viable but non-culturable state are not covered. Although HT-qPCR provided a limited coverage of the whole population in cheese, it has the ability to produce a more comprehensive and accurate evaluation with regard

Fig. 5 Comparison of HT-qPCR and 16S rRNA gene amplicon sequencing (NGS) data with corrections for *Lactiplantibacillus plantarum* group species and 16S rRNA gene copy normalization. Relative species compositions of the samples measured by HT-qPCR and NGS The samples are sorted and clustered according to the UPGMA linkage based on the Bray-Curtis dissimilarity. The relative abundance of the 15 species/groups detected by both methods are depicted with the species name; the other taxa were summarized as other species

to the abundance of the selected bacterial species than previous culture-dependent and qualitative culture-independent approaches. On one hand, precise identification at the species level can be achieved, as with qualitative culture-independent methods, and on the other hand, absolute abundances can be measured.

Other approaches currently in use include comparisons between different protocols, sequencers, and analysis pipelines for evaluating new protocols or benchmarking [3, 79, 80]. These approaches address a variety of additional parameters, such as the performance of sequencing platforms, the influence of primer choice and library

preparation protocols, as well as data analysis methods. These parameters were not within the scope of our study. Furthermore, there are aspects that are difficult to assess with these approaches, such as the influence of strain-specific variance in the number of copies of the rRNA operon or the species-specific PCR-associated bias [81]. Therefore, qPCR was particularly useful as a reference, because the influence of primer bias in qPCR reactions can be estimated accurately by the efficiency of the PCR reaction, and the variance of strain-specific 16S rRNA gene copy number per genome was bypassed by the selection of single copy gene targets.

A very detailed analysis is possible in our case because the microbial communities in the cheese core are quite well studied and shaped by the harsh and strictly controlled conditions during manufacturing and ripening [59, 82]. The HT-qPCR system used in this study was designed to cover frequent and abundant bacterial species in cooked, hard or semi-hard cheese with washed rinds made from raw milk. The data from 16S rRNA gene amplicon sequencing experiments was used to select appropriate target species to cover a high proportion of the most frequent and abundant bacterial community members. However, for other cheese types such as soft cheese varieties, the selection of the targets may be adapted. Likewise, the application of HT-qPCR to other fermented foods would be feasible. Since fermented foods such as sauerkraut or kimchi contain a core microbiota that partly overlaps with that of the cheese studied here [83, 84]. Nevertheless, this would most likely also require the development of some additional primer systems for species that are characteristic for the respective fermented food. For other, more complex ecosystems such as soil or the microbiota of specific human body sites, a comparison with the HT-qPCR approach is currently more challenging due to the high number of additional yet poorly defined taxa [85, 86].

Despite the identified bias for several species and the differences in 16S rRNA gene copy numbers, there was overall high agreement in relative abundances for the 15 species studied. This was an encouraging finding for the use of the NGS approach to study the microbiota of cheese.

A known, but sometimes neglected limitation of NGS approaches is that relative abundances are of limited use without knowledge of the absolute total abundance. Consequently, interpretation of the results may be challenging, especially for differential analyses or comparisons between samples with widely varying sampling or sequencing depths [19, 79, 87]. This is inherent in the method's principle because in sequencing experiments, the number of counts does not reflect the underlying absolute number of molecules in the sample, but rather the ratio of counts per OTU or ASV multiplied by sequencing depth [87]. For the analysis of cheese and milk related samples this information can be relevant as illustrated in the following concrete example. First, one can consider a scenario in which two LAB species, e.g. *Lc. lactis* and *L. fermentum*, are present in a natural whey culture at relative abundances of 90 and 10%, respectively, which corresponds to an absolute abundance of for instance 10^7 and 10^6 copies/ml. After incubation, the ratio detected by NGS is 50 and 50%. This could be due to the growth of *L. fermentum* to 10^7 copies/ml or to the reduction in the number of *Lc. lactis* to 10^6 copies/ml

due to autolysis or phage infection (while the growth of *L. fermentum* has stagnated). However, without knowing the total number of bacteria, distinguishing between the two scenarios is statistically more challenging and less accurate. In contrast, by using qPCR, we can measure the absolute abundance of the species directly without knowing the total number of bacteria.

As discussed above, we identified some flaws in taxa prediction accuracy using the selected NGS approach. Several studies have shown that the accurate identification of taxa depends on various factors such as the selected 16S rRNA gene region, read length, selected primers, sequencing platform, bioinformatics tools and reference databases [4, 7–9, 80, 88–90]. The approach for the bioinformatic analysis used in this study was already optimized by using an ecosystem-specific and manually curated reference database and bioinformatics algorithms with solid performance according to recent benchmarking studies [7, 89]. In the case of the already extensively discussed failure to differentiate the species of the *L. plantarum* group, the underlying cause was the high similarity of the 16S rRNA gene for these species. An improved prediction might only be achieved by longer reads or the selection of primers for a different target gene.

For NGS approaches, rarefaction curves are used to assess whether the sequencing depth is appropriate and if rare species/sequences could be identified with increasing sequencing depth [4]. However, the success to identify rare taxa and problems with a low sensitivity for certain taxa are often not only determined by the sequencing depth, the careful selection of the primers for the specific community under study is also important [90]. The low sensitivity for *P. pentosaceus* observed in the present study also indicates that species-specific PCR primer bias decreases the sensitivity significantly, even with an appropriate sequencing depth according to the saturation of the rarefaction curves (data not shown).

While qPCR is a well-established method with little potential for development beyond HT-qPCR, NGS is still experiencing a fast development. In this study, we applied single-end amplicon-based sequencing for the V1–V2 region of the rRNA gene with one analysis pipeline, including DAIRYdb and IDTAXA. Despite the limitations of NGS targeting only a small region, species annotation can be achieved thanks to highly curated databases [9]. In the future, amplicon free targeted sequencing by Nanopore can further improve the accuracy of NGS reducing the biases caused by preferential bindings of universal primers and allowing the sequencing of longer regions, such as the full 16S rRNA gene [1]. However, we also see future applications for HT-qPCR systems for the quantification of bacterial species in complex communities,

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 15 of 18

such as those found in other fermented foods. The modular design of HT-qPCR and the possibility of the fully automated primer design pipeline SpeciesPrimer [91] strongly simplify the challenging process of primer design for a high-throughput system and we believe will facilitate the adaptation of HT-qPCR to ecosystems other than cheese.

Conclusions

HT-qPCR and 16S rRNA gene amplicon sequencing provided highly comparable results for the qualitative and (semi-)quantitative characterization of bacterial communities in cheese. We have pointed out a number of differences and biases in measurements for several of the bacterial species included in this study. While the species assignments of most ASVs has been confirmed by HT-qPCR, we have also identified challenges in distinguishing *L. plantarum* from *L. pentosus* and in the correct assignment of *L. paraplantarum* based on the two variable regions of the 16S rRNA gene. Further, the different efficiencies for the measurement of several bacterial species were examined, and a potential PCR primer bias was identified.

We have highlighted the potential of NGS and HT-qPCR as complementary methods for both exploratory and screening purposes. NGS can be used to provide an overview of the microbial community, providing potential targets of interest for qPCR assay development, particularly in less known samples/environments. In return, qPCR can confirm species assignments, provide absolute quantitative data to better estimate the proportions of the bacterial composition, and draw attention to potential biases. HT-qPCR can then be used for more routine screening in environments with known bacterial composition.

Here, we demonstrated the application of NGS and HT-qPCR for the study of microbial communities in cheese and showed that the results were in substantial agreement. However, this approach may also be interesting for the study of the microbiota in other well-defined ecosystems.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-022-02451-y.

> **Additional file 1.**
> **Additional file 2.**
> **Additional file 3.**
> **Additional file 4.**

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Agroscope, Schwarzenburgstrasse 161, CH-3003 Bern, Switzerland. [2] Laboratory of Microbiology, University of Neuchâtel, Emile-Argand 11, CH-2000 Neuchâtel, Switzerland. [3] Department of Biomedicine, Applied Microbiology Research, University of Basel, Basel, Switzerland. [4] Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland. [5] Swiss Institute for Bioinformatics, Basel, Switzerland.

References
1. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22(1):178–93. https://doi.org/10.1093/bib/bbz155.
2. Boers SA, Jansen R, Hays JP. Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. Eur J Clin Microbiol Infect Dis. 2019;38(6):1059–70. https://doi.org/10.1007/s10096-019-03520-3.
3. Xue Z, Kable ME, Marco ML. Impact of DNA sequencing and analysis methods on 16S rRNA gene bacterial community analysis of dairy products. mSphere. 2018;3(5). https://doi.org/10.1128/mSphere.00410-18.
4. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010;38(22):e200. https://doi.org/10.1093/nar/gkq873.
5. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ. 2018;6:e4652. https://doi.org/10.7717/peerj.4652.
6. Bukin YS, Galachyants YP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci Data. 2019;6(1):190007. https://doi.org/10.1038/sdata.2019.7.
7. Prodan A, Tremaroli V, Brolin H, Zwinderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon

sequencing. PLoS One. 2020;15(1):e0227434. https://doi.org/10.1371/journal.pone.0227434.

8. Dueholm MS, Andersen KS, McIlroy SJ, Kristensen JM, Yashiro E, Karst SM, et al. Generation of comprehensive ecosystem-specific reference databases with species-level resolution by high-throughput full-length 16S rRNA gene sequencing and automated taxonomy assignment (AutoTax). mBio. 2020;11(5). https://doi.org/10.1128/mBio.01557-20.

9. Meola M, Rifa E, Shani N, Delbes C, Berthoud H, Chassard C. DAIRYdb: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. BMC Genomics. 2019;20(1):560. https://doi.org/10.1186/s12864-019-5914-8.

10. Ritari J, Salojarvi J, Lahti L, de Vos WM. Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. BMC Genomics. 2015;16:1056. https://doi.org/10.1186/s12864-015-2265-y.

11. Daisley BA, Reid G. BEExact: a metataxonomic database tool for high-resolution inference of bee-associated microbial communities. mSystems. 2021;6(2). https://doi.org/10.1128/mSystems.00082-21.

12. Gupta S, Mortensen MS, Schjørring S, Trivedi U, Vestergaard G, Stokholm J, et al. Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. Commun Biol. 2019;2(1):1–7. https://doi.org/10.1038/s42003-019-0540-1.

13. Poretsky R, Rodriguez RL, Luo C, Tsementzi D, Konstantinidis KT. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One. 2014;9(4):e93827. https://doi.org/10.1371/journal.pone.0093827.

14. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front Microbiol. 2016;7:459. https://doi.org/10.3389/fmicb.2016.00459.

15. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016;469(4):967–77. https://doi.org/10.1016/j.bbrc.2015.12.083.

16. Mas-Lloret J, Obon-Santacana M, Ibanez-Sanz G, Guino E, Pato ML, Rodriguez-Moranta F, et al. Gut microbiome diversity detected by high-coverage 16S and shotgun sequencing of paired stool and colon sample. Sci Data. 2020;7(1):92. https://doi.org/10.1038/s41597-020-0427-5.

17. Tessler M, Neumann JS, Afshinnekoo E, Pineda M, Hersch R, Velho LFM, et al. Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci Rep. 2017;7(1):6589. https://doi.org/10.1038/s41598-017-06665-3.

18. Giraffa G. Studying the dynamics of microbial populations during food fermentation. FEMS Microbiol Rev. 2004;28(2):251–60. https://doi.org/10.1016/j.femsre.2003.10.005.

19. Props R, Kerckhof FM, Rubbens P, De Vrieze J, Hernandez Sanabria E, Waegeman W, et al. Absolute quantification of microbial taxon abundances. ISME J. 2017;11(2):584–7. https://doi.org/10.1038/ismej.2016.117.

20. Kleyer H, Tecon R, Or D. Resolving species level changes in a representative soil bacterial community using microfluidic quantitative PCR. Front Microbiol. 2017;8(2017). https://doi.org/10.3389/fmicb.2017.02017.

21. Crane SL, van Dorst J, Hose GC, King CK, Ferrari BC. Microfluidic qPCR enables high throughput quantification of microbial functional genes but requires strict curation of primers. Front Environ Sci 2018;6; doi: https://doi.org/10.3389/fenvs.2018.00145.

22. Ishii S, Segawa T, Okabe S. Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. Appl Environ Microbiol. 2013;79(9):2891–8. https://doi.org/10.1128/aem.00205-13.

23. Hermann-Bank ML, Skovgaard K, Stockmarr A, Larsen N, Mølbak L. The Gut Microbiotassay: a high-throughput qPCR approach combinable with next generation sequencing to study gut microbial diversity. BMC Genomics. 2013;14(1):788. https://doi.org/10.1186/1471-2164-14-788.

24. Muhammed MK, Krych L, Nielsen DS, Vogensen FK. A high-throughput qPCR system for simultaneous quantitative detection of dairy *Lactococcus lactis* and *Leuconostoc* bacteriophages. PLoS One. 2017;12(3):e0174223. https://doi.org/10.1371/journal.pone.0174223.

25. Postollec F, Falentin H, Pavan S, Combrisson J, Sohier D. Recent advances in quantitative PCR (qPCR) applications in food microbiology. Food Microbiol. 2011;28(5):848–61. https://doi.org/10.1016/j.fm.2011.02.008.

26. Cao Y, Fanning S, Proos S, Jordan K, Srikumar S. A review on the applications of next generation sequencing technologies as applied to

27. Dreier M, Berthoud H, Shani N, Wechsler D, Junier P. Development of a high-throughput microfluidic qPCR system for the quantitative determination of quality-relevant bacteria in cheese. Front Microbiol. 2021;11:619166. https://doi.org/10.3389/fmicb.2020.619166.

28. Quigley L, O'Sullivan DJ, Daly D, O'Sullivan O, Burdikova Z, Vana R, et al. Thermus and the pink discoloration defect in cheese. mSystems. 2016;1(3). https://doi.org/10.1128/mSystems.00023-16.

29. Egger L, Ménard O, Abbühl L, Duerr D, Stoffers H, Berthoud H, et al. Higher microbial diversity in raw than in pasteurized milk Raclette-type cheese enhances peptide and metabolite diversity after *in vitro* digestion. Food Chem. 2021;340:128154. https://doi.org/10.1016/j.foodchem.2020.128154.

30. Xue J, Yang Y, Wang Z, Guo Y, Shao Y. Bacterial diversity in Chinese Rushan cheese from different geographical origins. Front Microbiol. 2018;9:1920. https://doi.org/10.3389/fmicb.2018.01920.

31. Raclette du Valais AOP. Raclette du Valais AOP. https://www.raclette-du-valais.ch (2021). Accessed 29 May 2021.

32. Eugster-Meier E, Fröhlich-Wyder M-T, Jakob E, Wechsler D, Morales MBL, Licitra G, et al. Semi-hard cheeses. In: Papademas P, Bintsis T, editors. Global Cheesemaking Technology: John Wiley & Sons, Ltd; 2017. p. 247–300.

33. Fluidigm. Gene expression with the 192.24 IFC using delta gene assays (100-7222 C1). https://www.fluidigm.com/binaries/content/documents/fluidigm/resources/192.24-ge-delta-gene-fast-std-qr-100-7222/192.24-ge-delta-gene-fast-std-qr-100-7222/fluidigm%3Afile (2015). Accessed 23 Oct 2020.

34. Fluidigm. Real-time PCR analysis user guide (68000088). https://www.fluidigm.com/binaries/content/documents/fluidigm/search/hippo%3Aresultset/real-time-pcr-analysis-ug-68000088/fluidigm%3Afile (2020). Accessed 23 Oct 2020.

35. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3. https://doi.org/10.1038/nmeth.3869.

36. Murali A, Bhargava A, Wright ES. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome. 2018;6(1):140. https://doi.org/10.1186/s40168-018-0521-5.

37. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217. https://doi.org/10.1371/journal.pone.0061217.

38. R Core Team. R. A language and environment for statistical computing. R Foundation for Statistical. Computing. 2020.

39. Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. 2015;43(0):D593–8. https://doi.org/10.1093/nar/gku1201.

40. Kluyver T, Ragan-Kelley B, Pérez F, Bussonnier M, Frederic J, Hamrick J, et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, editors. Positioning and Power in Academic Publishing: Players, Agents and Agendas: IOS Press; 2016. p. 87–90.

41. Perez F, Granger BE. IPython: a system for interactive scientific computing. Comput Scie Eng. 2007;9(3):21–9. https://doi.org/10.1109/MCSE.2007.53.

42. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2.

43. Michael Waskom, Maoz Gelbart, Olga Botvinnik, Joel Ostblom, Paul Hobson, Saulius Lukauskas, et al. mwaskom/seaborn: v0.11.1 (December 2020). Zenodo. 2020. https://doi.org/10.5281/zenodo.4379347.

44. Jeff Reback, Wes McKinney, jbrockmendel, Joris Van den Bossche, Tom Augspurger, Phillip Cloud, et al. pandas-dev/pandas: Pandas 1.2.3. Zenodo. 2021. https://doi.org/10.5281/zenodo.4572994.

45. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2.

46. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9(3):90–5. https://doi.org/10.1109/MCSE.2007.55.

Dreier *et al. BMC Microbiology* (2022) 22:48

Page 17 of 18

47. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference; 2010.

48. McLaren MR, Willis AD, Callahan BJ. Consistent and correctable bias in metagenomic sequencing experiments. Elife. 2019;8; doi: https://doi.org/10.7554/eLife.46923.

49. Matthias Dreier, Meola Marco, Hélène Berthoud, Noam Shani, Daniel Wechsler, Pilar Junier. biologger/htqpcr_ngs_data: release v1.0 (Version v1.0.0). Zenodo 2021. https://doi.org/10.5281/zenodo.4813080.

50. Seemann T. Barrnap 0.9: rapid ribosomal RNA prediction. https://github.com/tseemann/barrnap (2018). Accessed 23 Oct 2020.

51. Löytynoja A. Phylogeny-aware alignment with PRANK. Methods Mol Biol (Clifton, NJ). 2014;1079:155–70. https://doi.org/10.1007/978-1-62703-646-7_10.

52. Meola M, Rifa E. DAIRYdb v1.2.4: a manually curated reference database for improved taxonomy annotation of 16S rRNA gene sequences from dairy products. https://github.com/marcomeola/DAIRYdb/blob/master/DAIRYdb_v1.2.4_20200604/DAIRYdb_v1.2.4_20200604_blast/DAIRYdb_v1.2.4_20200604_blast.tgz (2020). Accessed 8 Dec 2020.

53. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90. https://doi.org/10.1093/bioinformatics/btl446.

54. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–W9. https://doi.org/10.1093/nar/gkz239.

55. Zago M, Bardelli T, Rossetti L, Nazzicari N, Carminati D, Galli A, et al. Evaluation of bacterial communities of Grana Padano cheese by DNA metabarcoding and DNA fingerprinting analysis. Food Microbiol. 2021;93:103613. https://doi.org/10.1016/j.fm.2020.103613.

56. Bertani G, Levante A, Lazzi C, Bottari B, Gatti M, Neviani E. Dynamics of a natural bacterial community under technological and environmental pressures: The case of natural whey starter for Parmigiano Reggiano cheese. Food Res Int. 2020;129:108860. https://doi.org/10.1016/j.foodres.2019.108860.

57. Masoud W, Vogensen FK, Lillevang S, Abu Al-Soud W, Sorensen SJ, Jakobsen M. The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. Int J Food Microbiol. 2012;153(1-2, 192):–202. https://doi.org/10.1016/j.ijfoodmicro.2011.11.014.

58. Quigley L, O'Sullivan O, Stanton C, Beresford TP, Ross RP, Fitzgerald GF, et al. The complex microbiota of raw milk. FEMS Microbiol Rev. 2013;37(5):664–98. https://doi.org/10.1111/1574-6976.12030.

59. Nam JH, Cho YS, Rackerby B, Goddik L, Park SH. Shifts of microbiota during cheese production: impact on production and quality. Appl Microbiol Biotechnol. 2021;105(6):2307–18. https://doi.org/10.1007/s00253-021-11201-5.

60. Mangia NP, Fancello F, Deiana P. Microbiological characterization using combined culture dependent and independent approaches of Casizolu pasta filata cheese. J Appl Microbiol. 2016;120(2):329–45. https://doi.org/10.1111/jam.13001.

61. Watanabe K, Fujimoto J, Tomii Y, Sasamoto M, Makino H, Kudo Y, et al. *Lactobacillus kisonensis* sp. nov., *Lactobacillus otakiensis* sp. nov., *Lactobacillus rapi* sp. nov. and *Lactobacillus sunkii* sp. nov., heterofermentative species isolated from sunki, a traditional Japanese pickle. Int J Syst Evol Microbiol. 2009;59(Pt 4):754–60. https://doi.org/10.1099/ijs.0.004689-0.

62. Han X, Zhang LJ, Wu HY, Wu YF, Zhao SN. Investigation of microorganisms involved in kefir biofilm formation. Antonie Van Leeuwenhoek. 2018;111(12):2361–70. https://doi.org/10.1007/s10482-018-1125-6.

63. Ennahar S, Cai Y, Fujita Y. Phylogenetic diversity of lactic acid bacteria associated with paddy rice silage as determined by 16S ribosomal DNA analysis. Appl Environ Microbiol. 2003;69(1):444–51. https://doi.org/10.1128/aem.69.1.444-451.2003.

64. Pogacic T, Mancini A, Santarelli M, Bottari B, Lazzi C, Neviani E, et al. Diversity and dynamic of lactic acid bacteria strains during aging of a long ripened hard cheese produced from raw milk and undefined natural starter. Food Microbiol. 2013;36(2):207–15. https://doi.org/10.1016/j.fm.2013.05.009.

65. De Bruyne K, Camu N, De Vuyst L, Vandamme P. *Lactobacillus fabifermentans* sp. nov. and *Lactobacillus cacaonum* sp. nov., isolated from Ghanaian cocoa fermentations. Int J Syst Evol Microbiol. 2009;59(Pt 1):7–12. https://doi.org/10.1099/ijs.0.001172-0.

66. Torriani S, Felis GE, Dellaglio F. Differentiation of *Lactobacillus plantarum*, *L. pentosus*, and *L. paraplantarum* by *recA* gene sequence analysis and multiplex PCR assay with *recA* gene-derived primers. Appl Environ Microbiol. 2001;67(8):3450–4. https://doi.org/10.1128/aem.67.8.3450-3454.2001.

67. Zhang F, Wang Z, Lei F, Wang B, Jiang S, Peng Q, et al. Bacterial diversity in goat milk from the Guanzhong area of China. J Dairy Sci. 2017;100(10):7812–24. https://doi.org/10.3168/jds.2017-13244.

68. Starke R, Pylro VS, Morais DK. 16S rRNA gene copy number normalization does not provide more reliable conclusions in metataxonomic surveys. Microb Ecol. 2021;81(2):535–9. https://doi.org/10.1007/s00248-020-01586-7.

69. Louca S, Doebeli M, Parfrey LW. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome. 2018;6(1):41. https://doi.org/10.1186/s40168-018-0420-9.

70. Pillidge CJ, Sheehy LM, Shihata A, Pu Z-Y, Dobos M, Powell IB. Intragenomic 16S rRNA gene heterogeneity in *Lactococcus lactis* subsp. *cremoris*. Int Dairy J. 2009;19(4):222–7. https://doi.org/10.1016/j.idairyj.2008.11.004.

71. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. J Bacteriol. 2004;186(9):2629–35. https://doi.org/10.1128/jb.186.9.2629-2635.2004.

72. Quigley L, O'Sullivan O, Beresford TP, Ross RP, Fitzgerald GF, Cotter PD. High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. Appl Environ Microbiol. 2012;78(16):5717–23. https://doi.org/10.1128/AEM.00918-12.

73. Alegria A, Szczesny P, Mayo B, Bardowski J, Kowalczyk M. Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches. Appl Environ Microbiol. 2012;78(6):1890–8. https://doi.org/10.1128/AEM.06081-11.

74. Dolci P, De Filippis F, La Storia A, Ercolini D, Cocolin L. rRNA-based monitoring of the microbiota involved in Fontina PDO cheese production in relation to different stages of cow lactation. Int J Food Microbiol. 2014;185:127–35. https://doi.org/10.1016/j.ijfoodmicro.2014.05.021.

75. Ryssel M, Johansen P, Al-Soud WA, Sorensen S, Arneborg N, Jespersen L. Microbial diversity and dynamics throughout manufacturing and ripening of surface ripened semi-hard Danish Danbo cheeses investigated by culture-independent techniques. Int J Food Microbiol. 2015;215:124–30. https://doi.org/10.1016/j.ijfoodmicro.2015.09.012.

76. Alessandria V, Ferrocino I, De Filippis F, Fontana M, Rantsiou K, Ercolini D, et al. Microbiota of an Italian Grana-like cheese during manufacture and ripening, unraveled by 16S rRNA-based approaches. Appl Environ Microbiol. 2016;82(13):3988–95. https://doi.org/10.1128/AEM.00999-16.

77. Cardinali F, Osimani A, Taccari M, Milanovic V, Garofalo C, Clementi F, et al. Impact of thistle rennet from *Carlina acanthifolia* All. subsp. *acanthifolia* on bacterial diversity and dynamics of a specialty Italian raw ewes' milk cheese. Int J Food Microbiol. 2017;255:7–16. https://doi.org/10.1016/j.ijfoodmicro.2017.05.018.

78. Ramezani M, Hosseini SM, Ferrocino I, Amoozegar MA, Cocolin L. Molecular investigation of bacterial communities during the manufacturing and ripening of semi-hard Iranian Liqvan cheese. Food Microbiol. 2017;66:64–71. https://doi.org/10.1016/j.fm.2017.03.019.

79. Llorens-Rico V, Vieira-Silva S, Goncalves PJ, Falony G, Raes J. Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. Nat Commun. 2021;12(1):3562. https://doi.org/10.1038/s41467-021-23821-6.

80. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics. 2016;17:55. https://doi.org/10.1186/s12864-015-2194-9.

81. Bonk F, Popp D, Harms H, Centler F. PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. J Microbiol Methods. 2018;153:139–47. https://doi.org/10.1016/j.mimet.2018.09.015.

82. De Filippis F, La Storia A, Stellato G, Gatti M, Ercolini D. A selected core microbiome drives the early stages of three popular italian cheese manufactures. PLoS One. 2014;9(2):e89680. https://doi.org/10.1371/journal.pone.0089680.

Dreier *et al. BMC Microbiology*   (2022) 22:48

Page 18 of 18

83. Plengvidhya V, Breidt F Jr, Lu Z, Fleming HP. DNA fingerprinting of lactic acid bacteria in sauerkraut fermentations. Appl Environ Microbiol. 2007;73(23):7697–702. https://doi.org/10.1128/AEM.01342-07.

84. Jung JY, Lee SH, Kim JM, Park MS, Bae JW, Hahn Y, et al. Metagenomic analysis of kimchi, a traditional Korean fermented food. Appl Environ Microbiol. 2011;77(7):2264–74. https://doi.org/10.1128/AEM.02157-10.

85. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551(7681):457–63. https://doi.org/10.1038/nature24621.

86. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59–65. https://doi.org/10.1038/nature08821.

87. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224. https://doi.org/10.3389/fmicb.2017.02224.

88. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):5029. https://doi.org/10.1038/s41467-019-13036-1.

89. Joos L, Beirinckx S, Haegeman A, Debode J, Vandecasteele B, Baeyen S, et al. Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. BMC Genomics. 2020;21(1):733. https://doi.org/10.1186/s12864-020-07126-4.

90. Zhan A, MacIsaac HJ. Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. Conserv Genet. 2014;16(3):513–22. https://doi.org/10.1007/s10592-014-0678-9.

91. Dreier M, Berthoud H, Shani N, Wechsler D, Junier P. SpeciesPrimer: a bioinformatics pipeline dedicated to the design of qPCR primers for the quantification of bacterial species. PeerJ. 2020;8:e8544. https://doi.org/10.7717/peerj.8544.

Publisher's Note