https://doi.org/10.35662/unine-thesis-2950

Characterization of the microbial community in raw milk cheese by high-throughput qPCR

A dissertation submitted to

UNIVERSITY OF NEUCHATEL

by

MATTHIAS DREIER

2022

Characterization of the microbial community in raw milk cheese by high-throughput qPCR

A dissertation presented to the Faculty of Science

University of Neuchatel

Institute of Biology

For the *Philosophiae doctor* degree

in Sciences

by

Matthias Dreier

Accepted by the Jury:

Prof. Pilar Junier, thesis director, University of Neuchatel

Prof. Philipp Engel, University of Lausanne

Prof. Ted Turlings, University of Neuchatel

Prof. Daniel Croll, University of Neuchatel

Dr. Hélène Berthoud, Agroscope

Dr. Noam Shani, Agroscope

Dr. Daniel Wechsler, Agroscope

Defended on August 18, 2021

University of Neuchatel

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Monsieur Matthias DREIER

Titre:

"Characterization of the microbial community in raw milk cheese by high-throughout qPCR"

sur le rapport des membres du jury composé comme suit:

- Prof. Pilar Junier, directrice de thèse, Université de Neuchâtel, Suisse
- Prof. Philipp Engel, Université de Lausanne, Suisse
- Prof. Ted Turlings, Université de Neuchâtel, Suisse
- Prof. Daniel Croll, Université de Neuchâtel, Suisse
- Dre Hélène Berthoud, Agroscope, Suisse
- Dr Noam Shani, Agroscope, Liebefeld, Suisse
- Dr Daniel Wechsler, Agroscope, Liebefeld, Suisse

Neuchâtel, le 3 mars 2022

Le Doyen, Prof. A. Bangerter

"*Gromit, that's it! Cheese! We'll go somewhere where there's cheese!*
Everybody knows the moon is made of cheese..."

Wallace

In: A Grand Day Out with Wallace and Gromit
directed and animated by Nick Park

Abstract

In recent decades, culture-independent methods have been extensively used for research and diagnostics in food microbiology. These methods allow a rapid and accurate determination of the microbial composition in a variety of foods. Recently, quantitative PCR (qPCR) and next-generation sequencing (NGS), especially 16S rRNA gene amplicon sequencing, were frequently used to investigate the microbial composition of cheese and its influence on cheese quality. However, these methods are relatively complex and expensive for a more widespread use, for instance, for the routine monitoring of cheese quality in cheese dairies. An advantage of qPCR over NGS is the quantification of the targeted bacterial species in cheese; however, there is also a trade-off in throughput and the bacterial population coverage. A few years ago, the development of new technologies in microfluidics enabled higher throughput for qPCR systems, making it possible to perform qPCR analyses on larger numbers of samples and assays with reduced effort. The aim of this thesis was to develop the basis for a high-throughput qPCR (HT-qPCR) system for the quantification of quality-relevant bacterial species in raw milk cheese, which would be applicable for research as well as for the diagnosis of quality problems.

For this purpose, an automated pipeline for the design of species-specific qPCR primers was developed and validated. In total, 23 new primer systems were designed targeting bacterial species commonly used in starters or present as part of the non-starter lactic acid bacteria in raw milk cheeses. The performance of these primer systems in combination with the HT-qPCR system were validated using DNA extracts from pure bacterial cultures and inoculated model cheeses. Further, the performance of HT-qPCR and 16S rRNA gene amplicon sequencing to determine the microbial composition of 21 Raclette du Valais cheeses was compared.

In this thesis, the basic concepts of the development of a HT-qPCR system for the quantification of some of the most important bacterial species with relevance to cheese quality were described. The application and usefulness for microbiological research of fermented foods was demonstrated. A potential application of this novel approach is the study of the bacterial community development during ripening. Further development of the modular system could eventually facilitate the rapid and economical screening of quality-relevant bacterial species in cheeses before and during ripening, thus ensuring continuous quality control during the manufacture of raw milk cheeses.

Keywords: qPCR primer design, bioinformatics, microbial community composition, microfluidics, cheese quality, cheese microbiome, fermented food, food microbiology

Résumé

Au cours des dernières décennies, les méthodes indépendantes de la culture ont été largement utilisées pour la recherche et le diagnostic en microbiologie alimentaire. Ces méthodes permettent une détermination rapide et précise de la composition microbienne d'une variété d'aliments. Récemment, la PCR quantitative (qPCR) et le séquençage de nouvelle génération (NGS), notamment le séquençage des amplicons du gène de l'ARNr 16S, ont été fréquemment utilisés pour étudier la composition microbienne du fromage et son influence sur la qualité du fromage. Cependant, ces méthodes sont relativement complexes et coûteuses pour une utilisation plus répandue, par exemple pour le contrôle de routine de la qualité du fromage dans les fromageries. Un avantage de la qPCR par rapport au NGS est la quantification des espèces bactériennes ciblées dans le fromage; cependant, il y a aussi un désavantage concernant le débit et la couverture de la population bactérienne. Il y a quelques années, le développement de nouvelles technologies dans le domaine de la microfluidique a permis d'augmenter le débit des systèmes qPCR, rendant possible la réalisation d'analyses qPCR sur un plus grand nombre d'échantillons et de tests avec un effort réduit. L'objectif de cette thèse était de développer les bases d'un système qPCR à haut débit (HT-qPCR) pour la quantification d'espèces bactériennes importantes pour la qualité des fromages au lait cru, qui serait applicable pour la recherche ainsi que pour le diagnostic des problèmes de qualité.

Dans ce but, un pipeline automatisé pour la conception d'amorces qPCR spécifiques aux espèces a été développé et validé. Au total, 23 nouveaux systèmes d'amorces ont été conçus pour cibler les espèces bactériennes couramment utilisées dans les starters ou les bactéries lactiques non starters présentes dans les fromages au lait cru. La performance de ces systèmes d'amorces en combinaison avec le système HT-qPCR a été validée en utilisant des extraits d'ADN de cultures bactériennes pures et de fromages modèles inoculés. De plus, les performances de la HT-qPCR et du séquençage des amplicons du gène de l'ARNr 16S ont été comparées pour déterminer la composition microbienne de 21 fromages Raclette du Valais.

Dans cette thèse, les concepts de base du développement d'un système HT-qPCR pour la quantification de certaines des espèces bactériennes les plus importantes pour la qualité du fromage ont été décrits. L'application et l'utilité de ce système pour la recherche microbiologique des aliments fermentés ont été démontrées. Une application potentielle de cette nouvelle approche serait l'étude du développement de la communauté bactérienne pendant la maturation. La poursuite du développement du système modulaire pourrait faciliter le dépistage rapide et économique des espèces bactériennes importantes pour la qualité des fromages avant et pendant la maturation, assurant ainsi un contrôle continu de la qualité pendant la fabrication des fromages au lait cru.

Mots clés: Conception de primer qPCR, bioinformatique, composition de la communauté microbienne, microfluidique, qualité du fromage, microbiome du fromage, aliments fermentés, microbiologie alimentaire

Acknowledgements

I thank Pilar Junier, my supervisor, for giving me the chance to collaborate with her laboratory. Thank you for your help, support and confidence and for giving me a lot of freedom in the realization of this work.

I thank Daniel Wechsler for the great support as supervisor and mentor over the last years. Thanks for always believing in me and cheering me up again. Thank you also for caring about my well-being and the reminder that there are other things in life apart from the thesis during times I was too ambitious.

I thank Noam Shani for his support with the countless strains of bacteria, the thorough review of my reports even when they came at short notice and the constructive discussions. I thank Hélène Berthoud for the wonderful idea of a high-throughput qPCR method for cheese, the critical examination of my work and the support during the last four years.

I thank Remo Schmidt for the accurate and precise way he gives feedback, for taking the time to review my manuscripts, and for supporting this thesis until the end.

I thank Daniel Marzohl, Nadine Sidler, Pascal Bätscher, Elvira Wagner, and Kotchanoot Srikham for their valuable technical help, especially for the cultivation of hundreds of strains, countless DNA extractions and qPCR experiments.

I thank Marco Meola and Vincent Somerville for the fruitful discussions and inspiration during and outside of work.

Thanks to the members of Pilar's lab for being open-minded, helpful and nice during my occasional visits.

I thank Aria Minder and Silvia Kobel for their support in performing the HT-qPCR analyses and all the people involved at the Center for Genetic Diversity (GDC) at ETH Zurich for creating such a valuable resource base.

I would like to thank the anonymous Reviewer 1 for the detailed review of the SpeciesPrimer manuscript, which has improved by far due to the critical and constructive comments. My thanks go to my friends and family for their continued support during these intense years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis outline

This thesis addresses the development of a new molecular biological method, namely high-throughput quantitative PCR (HT-qPCR), for the detection and enumeration of bacterial species of the microbial community in the cheese core. The developed HT-qPCR approach aims to help to better characterize and understand the quality-relevant bacterial communities in raw milk cheese. In the future, this molecular method could also be used for monitoring and contributing to improving cheese quality. The steps taken to achieve this goal are briefly described below.

This first chapter provides a general introduction to the bacterial communities in cheese core, their influence on the sensory and qualitative aspects, and an overview of the current methods used to study bacterial communities in cheese.

The second chapter describes the development and validation of a primer design pipeline used to design species-specific primers to detect bacteria in cheese. The pipeline was an important prerequisite for the later development of the HT-qPCR system.

The third chapter describes how the HT-qPCR system has been validated and provides proof of concept that the system works for the detection of 24 different bacterial species in cheese DNA samples.

The fourth chapter describes the application of HT-qPCR to 21 Raclette du Valais cheeses and compares the results with the results of a 16S rRNA gene amplicon sequencing approach to identify the strengths and limitations of these two approaches.

Finally, the fifth chapter follows a general discussion about the current methodology of cheese research and the impact of the HT-qPCR approach. In addition, an outlook on possible applications of HT-qPCR for cheese research and industry is given.

1.2 Raw milk

Raw milk cheeses are considered richer in flavor than cheeses made from pasteurized milk due to the beneficial impact of the microbial community in raw milk. Recent findings using *in vitro* digestion models have shown that, due to higher bacterial diversity, specific digestion-resistant peptides were increased in raw milk cheeses compared to pasteurized milk cheeses. The authors hypothesized that an increase in bioactive peptides in raw milk cheeses could benefit consumers (Egger et al., 2021). However, as a prerequisite, the raw milk used for cheese production has to be of good quality. Raw milk contains a diverse bacterial population dominated by members of the lactic acid bacteria (LAB) group. Several non-LAB bacteria, as well as yeasts and molds, are also found in milk (Quigley et al., 2013). Recent studies using molecular approaches have found that more than 100 bacterial species are present in raw milk (Quigley et al., 2011; Masoud et al., 2012). While some bacteria in raw milk contribute to the unique flavor of the resulting cheese, some species compromise the quality of the product and cause defects. For example, psychrotrophic bacteria, such as *Pseudomonas spp.* and *Acinetobacter spp.*, can grow to a significant portion of the population and cause milk spoilage even during cold storage (de Oliveira et al., 2015).

Moreover, the endospore former *Clostridium tyrobutyricum*, which can cause the late blowing of cheeses by butyric acid fermentation, is of great concern in milk used for cheese production (Klijn et al., 1995). Other bacteria frequently associated with microbial cheese quality problems include histamine-producing strains of *Lentilactobacillus parabuchneri* or dairy propionic acid bacteria (PAB) (Ascone et al., 2017; Berthoud et al., 2017; Turgay et al., 2018). Pathogens in raw milk that survive the manufacturing process are also highly relevant to cheese quality and food safety. Pathogens found in milk include *Listeria monocytogenes*, *Salmonella* spp., *Escherichia coli*, *Campylobacter* spp., and *Staphylococcus aureus* (Masoud et al., 2012).

1.3 Microbiology of cheese

1.3.1 Lactic acid bacteria (LAB)

LAB are a heterogeneous group of Gram-positive, catalase-negative, and non-spore-forming bacteria that produce lactate as the major metabolic product of carbohydrate fermentation. The genera often used in starter or adjunct cultures in cheese are *Lactococcus*, *Streptococcus*, *Lactobacillus*, *Leuconostoc,* and less frequently, *Enterococcus* (Fox et al., 2017c). A very important group of LAB for fermented foods is the *Lactobacillus* group because they are used in food production for millennia, and some strains are marketed as probiotics (Bernardeau et al., 2006). As for many other LAB, their taxonomic classification was initially based on parameters such as growth temperature and fermentation patterns. However, the genus *Lactobacillus* comprised a very diverse group of 261 species (Zheng et al., 2020), and new taxonomic classifications have been regularly proposed due to improved methods for 16S rRNA gene sequence and genome data analysis (Claesson et al., 2008; Sun et al., 2015; Salvetti et al., 2018). Recently, the reclassification of the genus *Lactobacillus* was realized, and a division into 25 genera was introduced (Zheng et al., 2020). The most important new genera related to the microbiota of cheese are listed in **Table 1.1**.

Table 1.1: Names and descriptions of new genera relevant for cheese as described by Zheng et al. (2020). Table from https://github.com/swuyts/lactotax.

Genus	Name description	Genus properties	Type species
Lacticaseibacillus	Lactobacilli related to cheese	Homofermentative, vancomycin-resistant; many species ferment pentoses and are resistant to oxidative stress; *L. casei* and related species have a nomadic lifestyle.	*L. casei*
Lactiplantibacillus	Lactobacilli related to plants	Homofermentative, vancomycin-resistant organisms with a nomadic lifestyle ferment various carbohydrates; most species metabolize phenolic acids by esterase, decarboxylase, and reductase activities. *Lactiplantibacillus plantarum* expresses pseudocatalase and nitrate reductase activities.	*L. plantarum*
Lactobacillus	Rod-shaped bacillus from milk	Homofermentative with strain-specific ability to ferment pentoses, thermophilic, vancomycin-sensitive, adapted to vertebrate or insect hosts.	*L. delbrueckii*
Latilactobacillus	Wide-spread lactobacilli	Homofermentative, vancomycin-resistant, mesophilic free-living, and environmental lactobacilli. Many strains are psychrotrophic and grow below 8°C.	*L. sakei*
Loigolactobacillus	(Food) spoiling lactobacilli	Homofermentative, vancomycin-resistant, mesophilic, or psychrotrophic organisms.	*L. coryniformis*
Lentilactobacillus	Slow (growing) lactobacilli	Heterofermentative, vancomycin-resistant, mesophilic, fermenting a broad spectrum of carbohydrates. Few lentilactobacilli appear to transition to a host-adapted lifestyle, but most are environmental or plant-associated. Most lentilactobacilli metabolize agmatine and convert lactate and/or diols.	*L. buchneri*

Table 1.1: Continued from previous page

Genus	Name description	Genus properties	Type species
Levilactobacillus	(Dough)-leavening lactobacilli	Heterofermentative, vancomycin-resistant, mesophilic or psychrotrophic, metabolize agmatine, environmental, or plant-associated lifestyle.	*L. brevis*
Limosilactobacillus	Slimy (biofilm-forming) lactobacilli	Heterofermentative, thermophilic, vancomycin-resistant with two exceptions, *Limosilactobacillus* species are vertebrate host-adapted and generally form exopolysaccharides from sucrose to support biofilm formation in the upper intestine of animals.	*L. fermentum*
Paucilactobacillus	Lactobacilli fermenting few carbohydrates	Heterofermentative, vancomycin-resistant, mesophilic, or psychrotrophic, aerotolerant, most strains ferment pentoses but not disaccharides.	*P. vaccinostercus*

1.3.2 Starter cultures

Starter cultures comprise different LAB and are responsible for the rapid acidification of milk at the beginning of cheese manufacture. Starter LAB are added to milk in large quantities, and in addition to acidification and lactose metabolism, they contribute to the taste of cheese through lipolysis and proteolysis during ripening. They can also contribute to the microbial safety of cheese by competing with pathogens and producing antimicrobial substances, such as bacteriocins (Arques et al., 2015; Mills et al., 2017).

Several types of starters differ in their optimum temperature and composition (Parente and Cogan, 2004). For mesophilic starters, mainly strains of *Lactococcus* (*Lc.*) *lactis* are used. Thermophilic cultures can comprise strains of *Streptococcus thermophilus* and Lactobacilli, such as *Lactobacillus delbrueckii* and *L. helveticus*. Further, there are mixed cultures with both mesophilic and thermophilic strains combined (Parente and Cogan, 2004). In traditional starters, such as whey cultures, the natural LAB contaminants of milk are selectively enriched by incubating whey or milk (Parente et al., 2016). Studies on traditional whey starter cultures for the production of Italian Grana Padano cheese identified *S. thermophilus*, *L. delbrueckii*, *L. helveticus,* and *Limosilactobacillus fermentum* as the dominant species (Rossetti et al., 2008; Santarelli et al., 2008). Today's starter cultures are derived from such natural starters, and certain strains are selected for defined mixed starter cultures. Alternatively, for undefined mixed starter cultures, established natural starters were selected without an in-depth characterization of the strains. There are ongoing efforts to isolate strains with positive characteristics, such as phage resistance or probiotic properties, from cheese, raw milk, plant material, and other sources to improve starter cultures.

1.3.3 Adjunct cultures

The main task of adjunct cultures is not the acidification of milk; they are selected to influence specific characteristics, such as flavor or eye formation. An example is the use of *Propionibacterium freudenreichii* in Swiss-type cheese, such as Emmental cheese. *P. freudenreichii* ferments lactate to propionic acid, acetate, and CO_2. The propionic acid gives the cheese a nutty and sweet flavor, and CO_2 supports the formation of typical eyes (Turgay et al., 2016). Selected strains of non-starter LAB (NSLAB) are sometimes used in adjunct cultures to support flavor formation or accelerate ripening, especially in cheeses made from pasteurized milk (Settanni and Moschetti, 2010).

1.3.4 Surface cultures

Surface cultures are used to support the development of surface microbiota on smear-ripened cheeses. The culture is added to the salt brine before smearing the cheese surface. These surface cultures can contain strain mixtures of different bacteria and yeasts, for example, *Debaryomyces hansenii* combined with *Brevibacterium linens* and *Arthrobacter* spp. (Agroscope, 2017).

1.3.5 Non-starter lactic acid bacteria (NSLAB)

NSLAB is the consortia of LAB, which is not part of the starter. NSLAB comprises mesophilic lactobacilli, pediococci, enterococci, and *Leuconostoc* spp. (Beresford and Williams, 2004). Their contribution to volatile flavor compound formation is mainly due to peptidolysis and the catabolism of amino acids (Settanni and Moschetti, 2010).

1.4 Cheese making

1.4.1 Manufacture

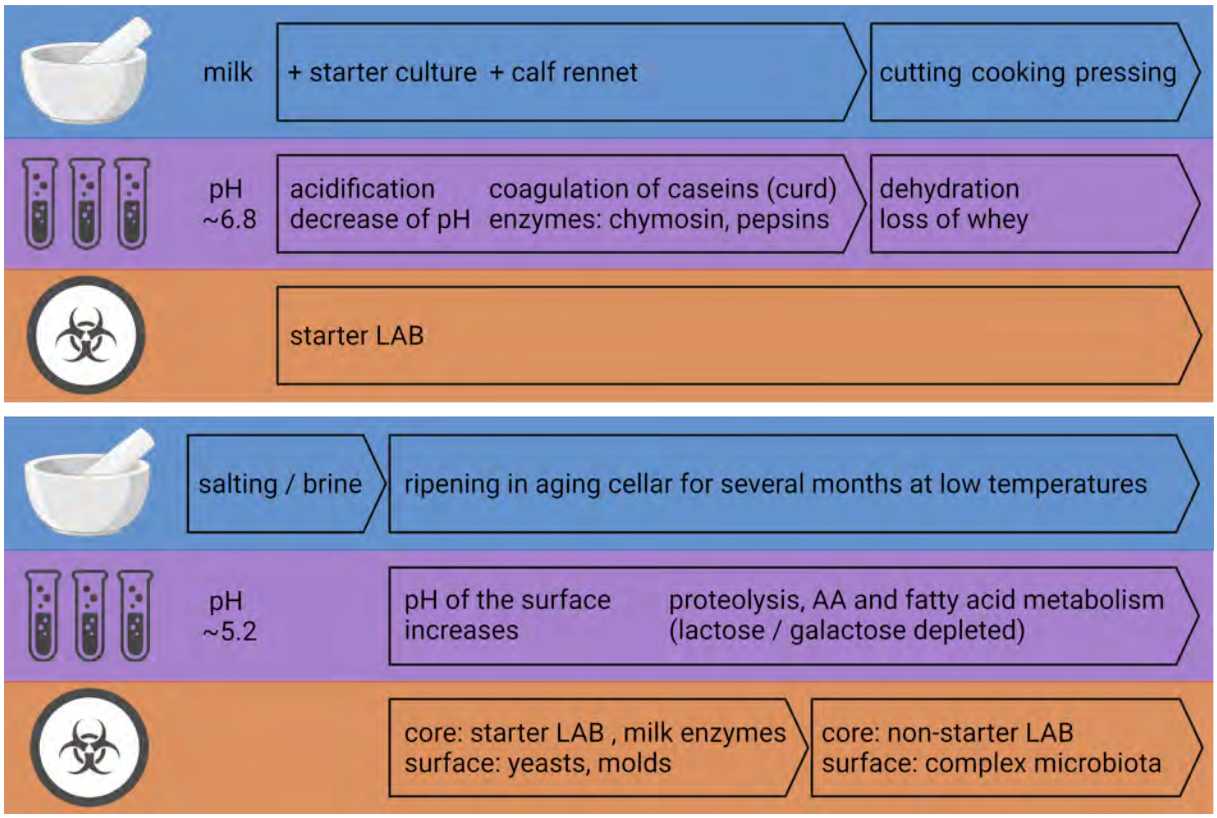


Figure 1.1: Simplified scheme of the steps for cheese manufacturing
The main steps during cheese manufacture are highlighted in blue, the chemical and biochemical processes in purple, and the predominant microbial groups in orange. The schematic view has been broken into two parts to facilitate reading. LAB: Lactic acid bacteria; AA: amino acid
Created with BioRender.com

The outline of cheese production for Swiss semi-hard and hard cheeses follows a certain pattern (**Figure 1.1**). Raw milk is preheated, and starter cultures are added. After the start of acidification by the starter LAB, rennet is added. The rennet enzymes, especially chymosin, support the coagulation of the caseins. The curd is then cut, the temperature is raised (scalding), and the whey is expelled from the curd (syneresis). The curd is then transferred to molds and pressed. The cheese loaves are then salted and stored on wooden shelves for ripening. Most Swiss semi-hard and hard cheese varieties are smear-ripened or dry-ripened. For smear-ripened varieties, cheese wheels are regularly turned and washed with fresh brine. The temperature and relative humidity in the aging room are strictly controlled. The duration of ripening varies from 3 to 12 months.

1.5 Biochemistry and microbiology of cheese ripening

1.5.1 Biochemistry of cheese ripening

Different chemical and biochemical reactions occur during cheese ripening. The combination of these reactions results in the flavor and texture characteristics of the final product. Three major biochemical pathways are important in cheese during ripening (**Figure 1.2**). First, the metabolism of residual lactose and the metabolism of lactate and citrate. Second, lipolysis and fatty acid metabolism, and the third is proteolysis and amino acid catabolism.

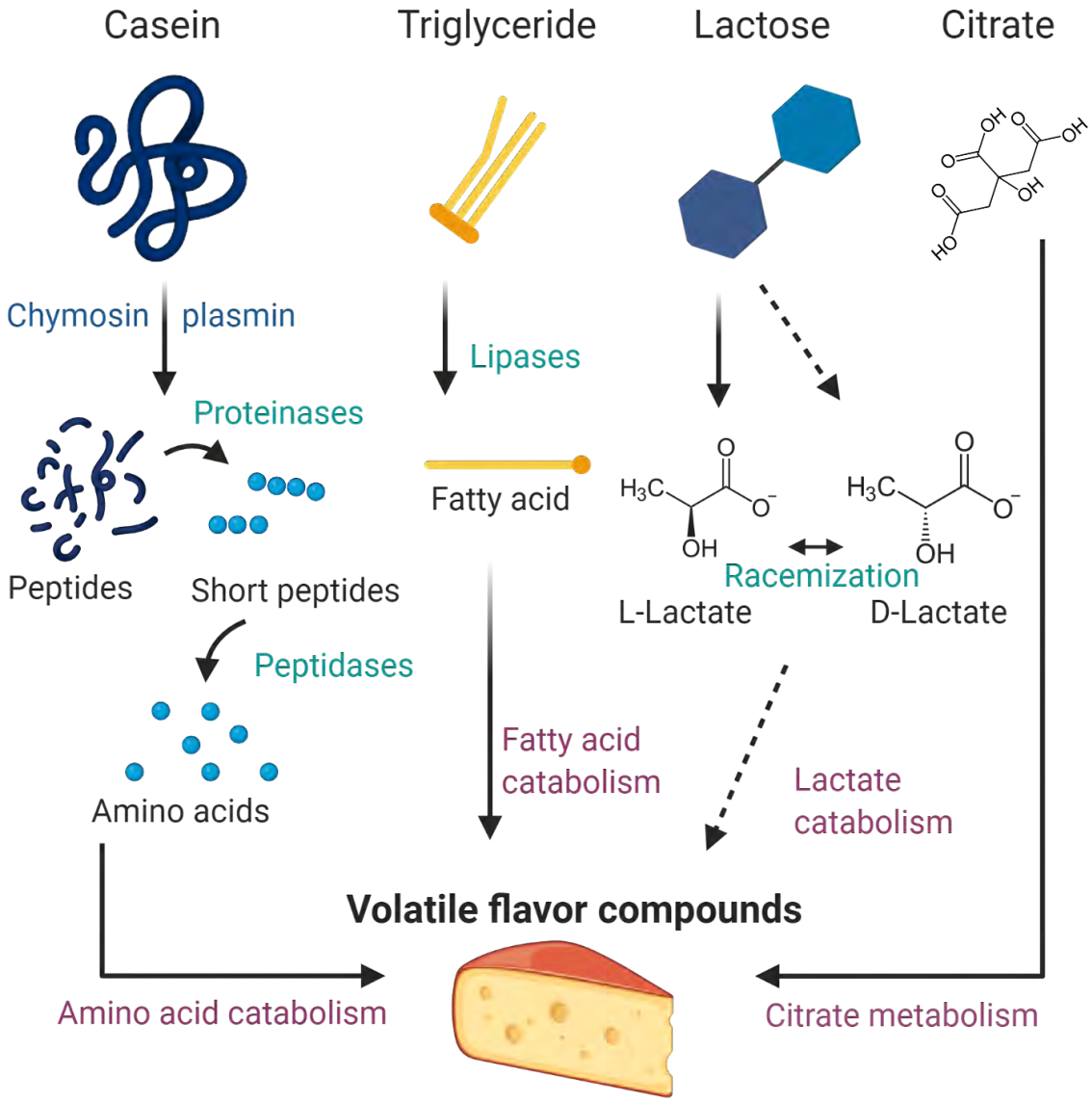


Figure 1.2: The major biochemical pathways in cheese during ripening
Adapted from (McSweeney, 2004), Created with BioRender.com

Metabolism of lactose, lactate, and citrate

Lactose is predominantly converted into lactate by the glycolysis of starter bacteria during cheese production and the early ripening stage. Residual lactose may be utilized as an energy source by NSLAB and spoilage bacteria, such as *Escherichia coli*. However, in Swiss cheeses, the presence of residual lactose in cheeses 24 hours

after production is expected to be the result of incomplete fermentation and is considered a defect, as this could contribute to the growth of undesirable bacteria during ripening.

Lactate has a substantial influence on the pH of cheese and, thereby, on the composition of the bacterial community and the metabolic activities of the bacteria during ripening (Linares et al., 2012; Fröhlich-Wyder et al., 2015). Certain species of NSLAB can metabolize lactate to acetate and CO_2 in the presence of O_2 (Thomas, 1987). However, this pathway is of minor importance in the cheese core. In surface-ripened cheeses, yeasts catabolize lactate in the early stages of ripening, which enables the formation of a selected surface microbiota by increasing the pH of the surface. Further, lactate can be catabolized anaerobically to acetate, CO_2, and 1,3-propanediol by *L. parabuchneri* (Oude Elferink et al., 2001), and it is also a substrate for butyric acid fermentation by *C. tyrobutyricum*, causing the cheese defect late blowing (Klijn et al., 1995). Furthermore, lactate is the substrate for propionic acid fermentation by propionic acid bacteria, which are important in Swiss-type cheeses. (Crow, 1986).

Citrate metabolism is a strain-level characteristic in LAB. In citrate-positive strains of lactococci, *Leuconostoc* spp., and lactobacilli, citrate is probably not used as an energy source, but in the presence of fermentable sugar, it is metabolized to diacetyl/acetoin, acetate, and CO_2 (Palles et al., 1998; McSweeney and Fox, 2004).

Lipolysis and fatty acid metabolism

In cheese, mainly short-and medium-chain free fatty acids (C4–C12) directly impact flavor due to a lower olfactory perception threshold compared to long fatty acids. Free fatty acids can also be precursors for other flavor compounds, such as esters, lactones, methyl ketones, or secondary alcohols (Thierry et al., 2017).

Lipases in cheese originate from milk, rennet, and the cheese microbiota. The indigenous milk lipoprotein lipase (LPL) contributes to lipolysis in raw milk cheeses, but it does not play a major role in cheeses made from pasteurized milk because of its low heat stability (McSweeney and Sousa, 2000; Deeth, 2006). In contrast to traditional rennet pastes, most rennet extracts nowadays do not show any lipolysis activity; thus, the main contribution to lipolysis in cheese originates from the esterases of the microorganisms. Starter LAB and NSLAB have a rather weak lipolytic activity but can contribute to lipolysis during long-time ripening. Among LAB, enterococci show the highest lipolytic activity (Tsakalidou et al., 1994), while PAB, such as *P. freudenreichii*, have a significantly higher lipolytic activity (Abeijón Mukdsi et al., 2014). Also, the surface microbiota of smear-ripened cheeses can contain strongly lipolytic bacteria, yeasts, and molds (Addis et al., 2001).

Proteolysis and amino acid catabolism

Caseins are the proteins that build the cheese matrix. Proteolysis, the successive degradation of caseins to smaller peptides and finally to amino acids, has an important influence on the texture and flavor of the cheese. Peptides and amino acids can directly impact flavor, while amino acids are also important precursors for the formation of aroma compounds.

Primary proteolysis of caseins is mostly due to added rennet (chymosin) and endogenous milk enzymes (plasmin). Starter bacteria harboring cell-envelope proteinases continue the degradation of caseins to oligopeptides. Intracellular peptidases released due to autolysis of starter bacteria degrade oligopeptides to shorter peptides and amino acids (Gatti et al., 1999). NSLAB are mainly involved in peptidolysis and the catabolism of amino acids. The most important cheese flavor compounds are derived from methionine, aromatic, and branched-chain amino acids (Upadhyay et al., 2004). The pathways of flavor formation from free amino acids have been reviewed in detail (Yvon and Rijnen, 2001; Ganesan and Weimer, 2017).

1.5.2 Microbiology of cheese ripening

There is a multitude of abiotic and biotic factors that influence the growth of microorganisms in cheese, for example, water activity (water availability), pH, salt concentration, temperature, reduction (-oxidation) potential, microbial interactions (competition, bacteriocins), and substrate availability. Together, these factors control microbial activity during ripening. In the early stages of ripening, the microbial community is dominated by starter LAB. They can grow to final numbers of up to 10^9 colony-forming units (cfu)/g of cheese within the first 24 hours. During ripening, starter bacteria undergo autolysis and release intracellular enzymes, such as peptidases and lipases, into the cheese matrix. These enzymes are of high importance for the development of the flavor and texture of cheese. Amongst others, autolysis was shown in *Lc. lactis* (O'Sullivan et al., 2000; Visweswaran et al., 2017), *L. helveticus,* and PAB (Valence et al., 1998; Treimo et al., 2006). However, viable starter LAB are also present at the end of cheese ripening, as shown by studies using reverse transcription quantitative PCR (RT-qPCR) (Falentin et al., 2012; Ruggirello et al., 2014). Ganesan et al. (2007) showed that under stress conditions, such as carbohydrate starvation, *Lc. lactis* cells change to a viable but non-culturable (VBNC) state.

The influence of viable starters during the later stages of ripening was possibly underestimated by classical culture-dependent methods (Ruggirello et al., 2018).

In the later stages of cheese ripening, NSLAB, mainly of (raw) milk origin, can grow and dominate the microbiota of the cheese. NSLAB are also present in cheeses made from pasteurized milk, although in lower amounts. NSLAB that survive in cheese are acid and salt tolerant and can adapt to environmental stresses and changing conditions during cheese ripening (Gatti et al., 2014). At the time of NSLAB growth, there is possibly only residual lactose present. In Swiss cheeses, such as Emmental, Gruyère, or Raclette no residual lactose is expected after the first day of ripening because the starter LAB rapidly metabolize all the lactose (McSweeney and Fox, 2004). Citrate, small peptides and amino acids, carbohydrates from glycomacropeptides of caseins, and glycoproteins of the milk fat globule membrane have been proposed as possible substrates for NSLAB growth (Beresford and Williams, 2004). Heterofermentative lactobacilli can possibly also metabolize pentoses derived from lysed starter bacteria (Rapposch et al., 1999). Despite the many efforts to study NSLAB in cheese, the details on the effects of their growth on flavor formation and cheese quality are still poorly understood.

On the surface of smear-ripened cheese, a microbiota develops whose composition is mainly determined by the natural microbiota of the ripening cellar by adding surface cultures and/or by transferring the microbiota from older to younger cheeses during smearing. For the latter, old cheeses with established surface microbiota are smeared with fresh brine, and the recovered brine is then used to smear the young cheeses. The complex surface microbiota has an advantageous influence on flavor development and is responsible for the characteristic color of the rind. Flavor development is associated with the production of volatile compounds by the lipolytic and proteolytic activity of the surface microbiota.

A first important step is the development of yeasts (e.g., *Debaryomyces hansenii*, *Kluyveromyces marxianus, and Geotrichum candidum*) on the surface. Yeasts metabolize lactate and produce alkaline metabolites, such as ammonia by deamination of amino acids. The deacidification of the surface enables the establishment of a salt-tolerant complex microbiota, which typically comprises staphylococci, *Arthrobacter* spp., *Brevibacterium* spp., *Corynebacterium* spp., *Microbacterium* spp., *Carnobacterium* spp., bacilli, and enterococci (Cogan et al., 2014; Wolfe et al., 2014). Also, various Gram-negative species, e.g., *Citrobacter* spp., *Proteus* spp. *Morganella* spp., and *Serratia* spp., may be found on the surface of smear-ripened cheeses. During ripening, the composition of the microbiota changes over time, and different species may be present successively (Larpin et al., 2006; Mounier et al., 2007; Cogan et al., 2014; Wolfe et al., 2014).

1.6 Cheese defects

1.6.1 Early gas defects

Excess gas production in cheese manifests as cracks, slits, and holes that impair consumers' acceptance of the cheese. Yeasts present in milk can cause early gas defects through the production of CO_2 from lactose or lactate (Fox et al., 2017b; Fox et al., 2017a). Coliforms from the genera *Enterobacter*, *Escherichia*, *Citrobacter,* and *Serratia* can be present in raw milk. These bacteria produce H_2 from formic acid as a by-product of lactose utilization (McSweeney, 2007). Other causative agents of early gas defects are sometimes citrate-positive lactococci or *Leuconostoc* spp., which produce CO_2 by lactose and citrate fermentation (McSweeney, 2007).

1.6.2 Late gas defects

The bacterium *C. tyrobutyricum* is the main cause of a cheese defect called late blowing, which is frequently observed in semi-hard and hard cheese varieties (Klijn et al., 1995). In addition, other members of the genus *Clostridium,* such as *C. butyricum*, *C. beijerinckii,* and *C. sporogenes,* are sometimes detected in cheeses with late blowing defects (Bassi et al., 2015). *C. tyrobutyricum* ferments lactate to butyric acid, CO_2, and H_2. High concentrations of butyric acid cause a taste defect, and the produced gas leads to the formation of big holes and the blowing of the cheese. Propionic acid bacteria (PAB) present in raw milk can have detrimental effects, changing the typical flavor and appearance (eyes) due to propionic acid and CO_2 formation. Facultative heterofermentative lactobacilli (FHL) originating from raw milk can also be responsible for gas defects in soft and semi-hard cheeses. These FHL, such as *L. fermentum*, *Latilactobacillus curvatus,* and *Levilactobacillus brevis,* can ferment residual lactose and produce CO_2 as a byproduct (Quiberoni et al., 2008; Porcellato and Skeie, 2016). Recently, a newly reported NSLAB species, *Paucilactobacillus wasatchensis,* was found to be involved in split formation in cheddar-style cheese. Interestingly the defect was only observed in combination with the starter *S. thermophilus*, which does not metabolize the galactose moiety of lactose. Accumulation of galactose was identified as a reason for the stimulation of growth of *Paucilactobacillus wasatchensis* and gas production (Ortakci et al., 2015).

1.6.3 Flavor and aroma defects

The flavor and aroma development by the cheese microbiota during ripening is a desired characteristic. However, if flavor compounds exceed a certain limit or the composition of flavor compounds is altered drastically, aroma and taste defects are observed.

Hydrophobic peptides often cause the bitter taste of cheese. Bitter hydrophobic peptides can accumulate during ripening by the increased proteolysis of caseins or by low starter peptidase activity (O'Sullivan et al., 2013). Mainly peptides from the C-terminal regions of β-casein and α_{s1}-casein are assumed to contribute to bitterness (Lemieux and Simard, 1991). Proteases that contribute to the release of bitter peptides are chymosin, from rennet in cheeses with low scalding temperatures, plasmin from milk, and cell envelope-associated proteinases (CEP) from starter LAB (Karametsi et al., 2014). Also, extracellular proteinases from *Pseudomonas* spp. in the milk can lead to an accumulation of bitter peptides in cheese (Lemieux and Simard, 1991).

Rancidity is a result of an excess of hydrolytic degradation of fatty acids. Secondary organisms, such as bacteria and fungi of the smear microbiota, can be very lipolytic. Natural rennet paste may contain potent lipases, such as pregastric esterase. Damage of the milk fat globule membrane increases the contribution of the milk LPL to lipolysis in raw milk cheese (McSweeney, 2007). Other flavor defects may result from the overproduction of volatile flavor compounds, such as ethyl esters, methyl alcohols, aldehydes, or acetaldehydes. Many of these defects are linked to starters and rarely arise because of careful starter selection in most cheese varieties (O'Sullivan et al., 2013).

1.6.4 Discoloration defects

Discoloration defects can arise from pigmented microbes forming visible spots (e.g., PAB) or from pigments secreted into the cheese matrix.

Methodological advances have recently made it possible to identify the cause of the cheese defect "pink discoloration". Quigley et al. (2016) used the 16S rRNA gene and shotgun metagenomic sequencing to compare cheeses with defects and control cheeses without defects. The only taxonomic group that differed significantly between the defect and control cheeses was the genus *Thermus*. Further investigation revealed that the production of a carotenoid by *Thermus thermophilus* appears to cause the pink discoloration defect. Cheeses with and without *T. thermophilus* were prepared, and an influence of the starter composition on the intensity of the discoloration was found. In the first experiment, the authors of the study tested a starter with *S. thermophilus* strains only and detected pink discoloration in the cheese. In a second experiment, the additional use of *L. helveticus* in the starter resulted in a more intense discoloration. However, the detailed mechanism by which lactobacilli contribute to an increase in pink coloration (i.e., whether they contribute to an increase in carotenoid production or influence other factors) was not investigated in this study.

1.6.5 Biogenic amines

Table 1.2: Microorganisms involved in the formation of biogenic amines in cheese

Biogenic amines (Amino acid substrate)	**Species**
Histamine (Histidine)	*L. parabuchneri*, *L. coryniformis*, *Morganella morganii*, *Debaryomyces hansenii*, *Geotrichum candidum*
Tyramine (Tyrosine)	*E. faecalis*, *E. faecium*, *E. durans*, *E. hirae*, *L. brevis*, *L. curvatus*, *Yarrowia lipolytica*
Putrescine (Arginine/Agmatine)	*L. curvatus*, *E. faecalis*, *E. hirae*, *Debaryomyces hansenii*, *Yarrowia lipolytica*, *Proteus* spp.
Cadaverine (Lysine)	*Shigella flexneri*, *Shigella sonnei*, *Escherichia coli*, *Hafnia* spp., *Salmonella* spp., *Vibrio* spp., *Bacillus* spp., *Clostridium* spp., *Listeria* spp., *Staphylococcus* spp.

Biogenic amines are amines with biological activity that can be formed during fermentations, e.g., during cheese ripening, by the decarboxylation of amino acids. The most common biogenic amines in cheese are histamine, tyramine, putrescine, and cadaverine, while beta-phenylethylamine and tryptamine are less common and usually present in lower concentrations. The ingestion of larger quantities of biogenic amines may cause toxic reactions. Normally, biogenic amines are degraded in the gut by amine oxidases. However, if this detoxification is disturbed, biogenic amines can cause severe health problems. Histamine toxicity can trigger neurological, gastrointestinal, and respiratory problems. Tyramine toxicological effects include headaches, migraines, and hypertension. Putrescine and cadaverine do not have direct toxic effects in concentrations usually found in cheese; however, they potentiate the effect of other biogenic amines, and they can form carcinogenic nitrosamines via a reaction with nitrite (Ladero et al., 2010a; Linares et al., 2011).

Biogenic amine formation in cheese is dependent on microorganisms with decarboxylase activity, the presence of the relevant amino acid substrate, and conditions favoring decarboxylation reactions. Different microorganisms of the cheese microbiota have been identified as potential biogenic amine producers; an overview is shown in **Table 1.2**. The ability to produce biogenic amines is often a strain level characteristic. Strains of Gram-negative bacteria of the family *Enterobacteriaceae* as *Hafnia alvei*, *Serratia liquefaciens*, *Escherichia coli*, *Citrobacter freundii,* or *Enterobacter* spp., have shown potential for biogenic amine formation (de las Rivas et al., 2006; Pircher et al., 2007). Also, yeasts and molds, such as *Debaryomyces hansenii* and *Yarrowia lipolytica*, which may be present on the surface of smear-ripened cheese varieties, were identified as potential biogenic amine producers (Gardini et al., 2006). However, the main group responsible for high amounts of biogenic amines found in the cheese core are the Gram-positive LAB. Starter LAB strains of *S. thermophilus* can produce histamine from histidine (Rossi et al., 2011) and strains of *Lc. lactis* were found to be able to produce putrescine (del Rio et al., 2015). Nowadays, commercial starter strains are screened for their decarboxylation capacity, so the production of biogenic amines by starter strains is very unlikely. Therefore, the focus on biogenic amine formation in cheese is on the group of NSLAB (Ladero et al., 2010a; Linares et al., 2012; Fröhlich-Wyder et al., 2013). Recently, *L. parabuchneri* was identified as one of the most potent producers of histamine in cheese (Diaz et al., 2016b; Ascone et al., 2017; Berthoud et al., 2017). Diaz et al. (2016a) investigated the potential of biogenic amine-producing bacteria isolated from dairy products to form biofilms on stainless steel. The results showed that most strains could adhere to stainless steel and could, therefore, be a possible source of contamination for milk and cheese throughout the entire cheese-making process from the farm to the cheese dairy.

1.7 Technical aspects of the methods discussed in the thesis

1.7.1 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was developed in the early 1980s for the amplification of specific DNA regions (Saiki et al., 1985; Saiki et al., 1988) and has since then been used for numerous applications (see Green and Sambrook, (2019) for a review). The main components of a PCR are a thermostable DNA polymerase, a primer pair (synthetic oligonucleotides), deoxynucleoside triphosphates (dNTPs), a buffer to maintain the pH (e.g., Tris-Cl), monovalent, and divalent cations (e.g., K^+ and Mg^{2+}) and template DNA. Repeated temperature cycles are used to perform the different steps of the PCR. In the first step, the temperature is raised for the denaturation of the DNA template (> 90°C). In a second step, the temperature is decreased to an optimized temperature (often between 55°C and 60°C) that facilitates annealing of the oligonucleotide primers to the denatured template. Subsequently, in the extension step, DNA synthesis is catalyzed by the DNA polymerase at temperatures between 55°C and 72°C. These steps are performed in a thermocycler and are usually repeated about 25–35 times.

1.7.2 PCR primer design

Primers are synthetic oligonucleotides that prime DNA synthesis by annealing to the complementary sequence of the denatured template DNA to form a short double-stranded sequence at which the DNA polymerase can initiate the extension of the DNA strand. The design of appropriate primers is one of the most important factors for the specificity and efficiency of a PCR. The most important factors for the design of specific primers and primer design using software have been extensively reviewed (Dieffenbach et al., 1993; Chuang et al., 2013).

There are some rules of thumb for the design of primer pairs that are briefly described below. The guidelines for primer design may vary for different applications and protocols and may deviate from these rules of thumb. The length of the primers should usually be between 16 and 28 nucleotides, and the guanine and cytosine (GC) content should be between 40% and 60%. Long GC-rich stretches, internally repeated, and self-complementary sequences should be avoided to lower the potential to form secondary structures, such as hairpin loops. The two primers

should not contain cross-complementary sequences, especially at the 3'-end, to reduce primer dimer formation. A GC clamp, the presence of G or C bases at the 3'-end of primers, can promote specific binding; however, more than three Gs or Cs in the last five bases of the 3'-end of primers should be avoided. The melting temperature of the primer (T_m) is the temperature at which half of the primer is dissociated and single-stranded; it indicates the stability of the template-primer DNA duplex. The T_m difference between primers of a primer pair should not differ more than 2°C–3°C. The T_m of a primer can be predicted using the thermodynamic nearest-neighbor model (Breslauer et al., 1986; SantaLucia, 1998; von Ahsen et al., 2001). The selected annealing temperature of the PCR is often about 3°C–5°C below the T_m of the primers. The appropriate selection of the annealing temperature is crucial because if the annealing temperature is too low non-specific annealing can be promoted; if it is too high, the annealing to the template DNA may be impaired.

There are several software tools available to help with the design and the *in silico* validation of primer properties, such as melting temperature calculation, estimation of dimers and hairpins, and primer specificity. For example, primer3 (Untergasser et al., 2012) is a tool for primer design and to calculate the melting temperatures, mfold (Zuker et al., 1999) allows for the evaluation of secondary structures, and BLAST (Altschul et al., 1990) can be used to evaluate primer's specificity. Primer-BLAST uses primer3 for primer design and BLAST for a subsequent specificity evaluation (Ye et al., 2012).

1.7.3 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) extends PCR by monitoring the amplification of the PCR products (amplicons) in real-time through the measurement of fluorescence. This real-time monitoring is achieved with specially developed thermocyclers with an integrated fluorimeter. There are several different fluorescence techniques available to assess the concentration of the amplicons. These include, for example, double-stranded DNA (dsDNA) specific dyes, such as SYBR Green I and EvaGreen (Mao et al., 2007), hydrolysis probes (TaqMan probes), hybridization probes, light-up probes, molecular beacon, Sunrise primer, or Scorpion primer (Wilhelm and Pingoud, 2003). In the following lines, the principle of qPCR using non-specific dsDNA dyes is explained. In addition to the abovementioned components of PCR, a dsDNA-binding dye is required for qPCR. The fluorescence is measured after each cycle, and the intensity of the measured fluorescence signal corresponds to the amount of DNA amplicons in the sample. The quantification cycle (Cq) is the cycle at which the fluorescence signal reaches a predetermined fluorescence threshold (**Figure 1.3A**). The initial quantity of the target sequence can be estimated based on the Cq values using a calibration curve, which is often constructed using ten-fold serial dilutions of standard samples with known concentrations. PCR is an exponential process, and if the efficiency is 100%, the number of target sequences in the reaction is doubled with each cycle. The amplification reaction can be expressed with the following equation, where: N is the number of amplicons, n is the cycle number, and E is the PCR efficiency.

$$N_n = N_0 * (1 + E)^n$$

If the PCR efficiency is 100%, the number of cycles between ten-fold dilutions is 3.322. In practice, the PCR efficiency is calculated from the slope (k) of the logarithm of the concentration/copies of the calibration curve from the ten-fold standard dilution series plotted on the x-axis and the corresponding Cq values on the y-axis (**Figure 1.3B**) (Bustin et al., 2009).

$$E = 10^{-(\frac{1}{k})} - 1$$

The concentration/number of copies for the samples can then be calculated from the Cq value using the following equation. The intercept (c) reflects the Cq value when one copy is theoretically detected.

$$y = kx + c$$

After the final cycle, a melting curve analysis is performed for qPCR with non-specific dsDNA dyes to assess that only the target amplicon is responsible for the measured fluorescence signal (**Figure 1.3C**). For the melting curve analysis, the temperature is increased stepwise, e.g., from 60°C to 95°C, while the fluorescence is constantly monitored. The decrease in the fluorescence signal due to dissociation of the dsDNA to single strands at a given melting temperature indicates whether the amplicon corresponds to the expected amplified product and allows the detection of multiple amplicons due to non-specific primer binding.

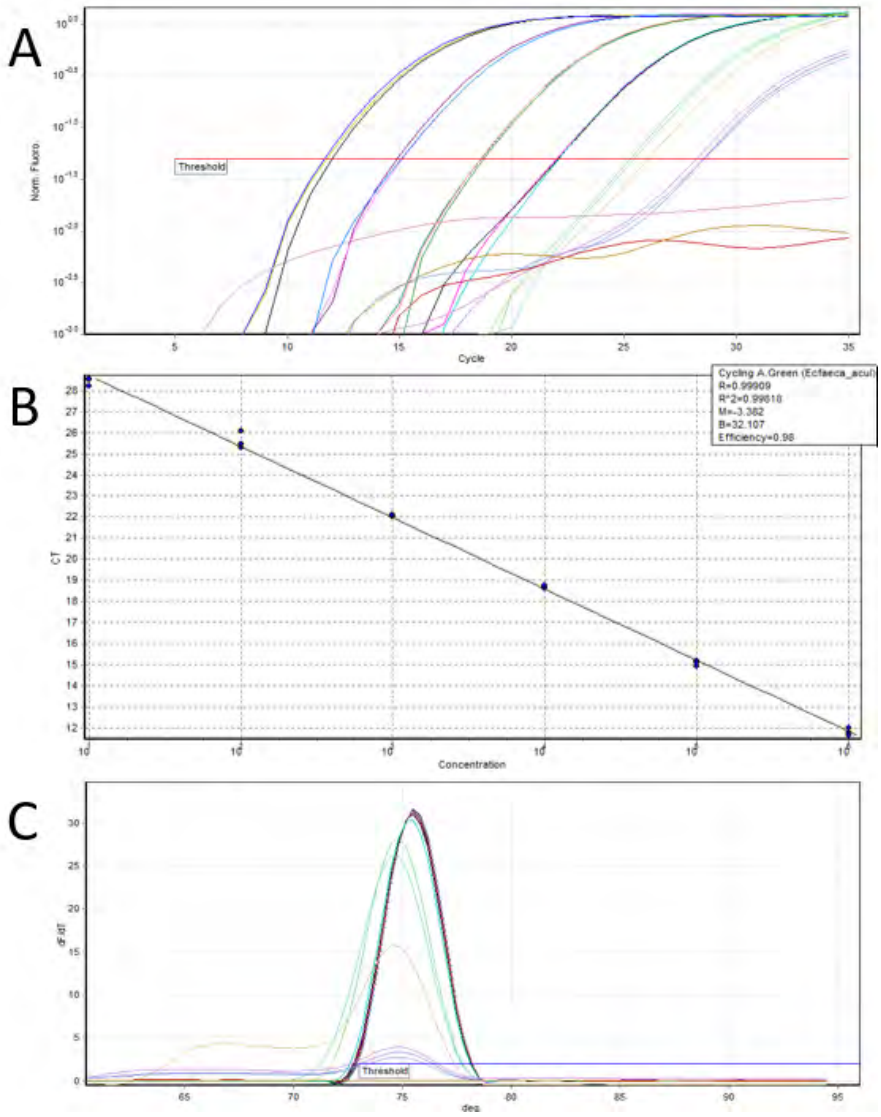


Figure 1.3: Example of an analysis of qPCR standard dilutions using Rotor-Gene 6000 Software
A) Fluorescence signals of ten-fold dilution series using type strain DNA with known concentration. B) Calibration curve with the parameters of the linear regression equation (correlation coefficient (R^2), slope (M), intercept (B), and PCR efficiency). C) Peaks of the melting curve analysis for the dilution series.

1.7.4 Next-generation and 16S rRNA gene amplicon sequencing

Since the invention of Sanger sequencing in the 1980s (Sanger et al., 1977), sequencing techniques have been continuously developed. Although the third generation of sequencing techniques is now available, the focus here is on the second generation, also called next-generation sequencing (NGS). The major advancement in this generation of sequencers was an enormous increase in throughput due to parallelization and the avoidance of electrophoresis-based sequence determination. The evolution of sequencing instruments, key technologies and applications were recently reviewed (Heather and Chain, 2016; Gupta and Gupta, 2020). Three frequently used next-generation sequencing technologies—454 pyrosequencing, Solexa/Illumina sequencing, and Ion Torrent—are briefly described below.

Pyrosequencing instruments manufactured by Roche/454 Life Sciences use water-in-oil emulsion PCR, in which DNA fragments ligated to adapter nucleotides are diluted to single-molecule concentration and hybridized to beads. On the beads in water-in-oil droplets, the single molecules are amplified to produce a clonal DNA population. A picotiter plate (flow-cell) that fits one bead per well and the enzymes required for sequencing is used to perform pyrosequencing. Pyrosequencing requires two enzymes, an ATP sulfurylase, to convert pyrophosphate into adenosine triphosphate (ATP), and luciferase, which emits light in proportion to the amount of

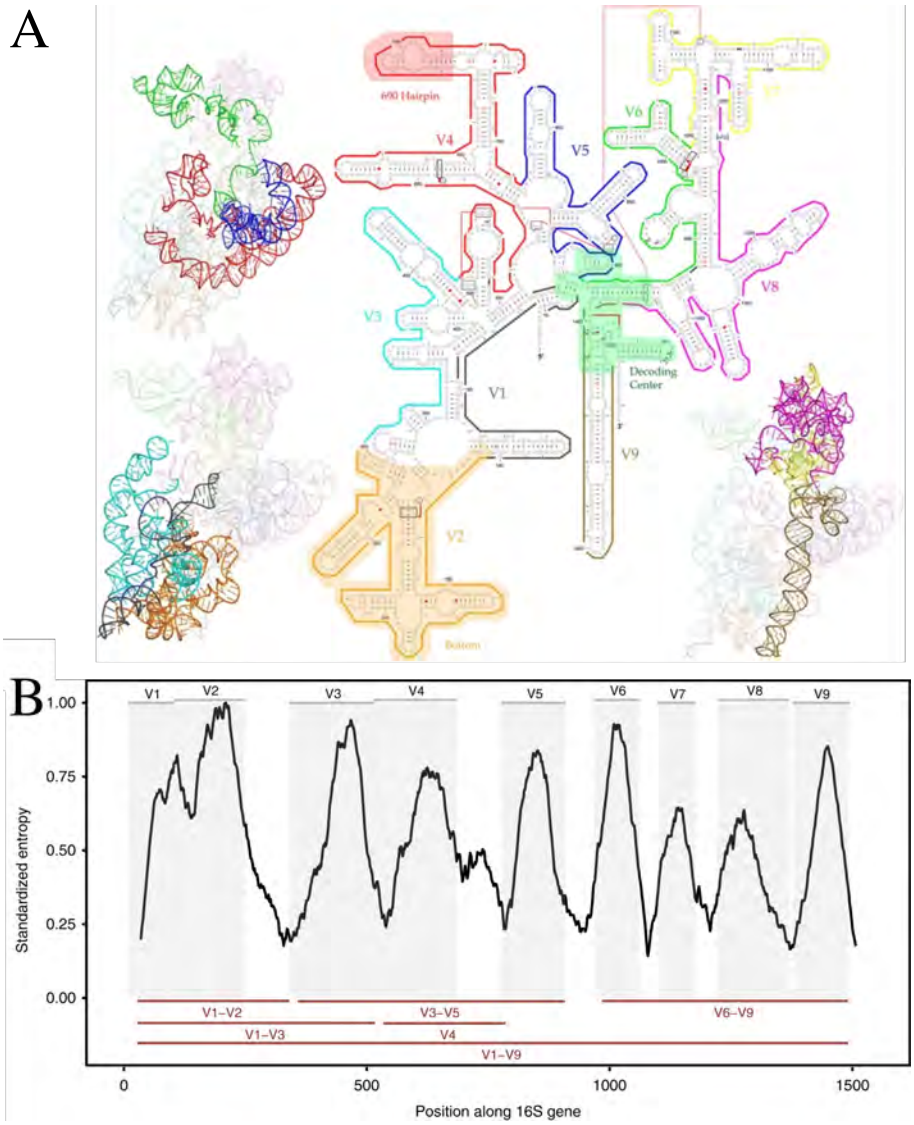


Figure 1.4: Structure of the 16S rRNA
A) The two-and three-dimensional structures of the 16S rRNA. The various regions are represented in different colors. Taken from (Yang et al., 2016). B) Shannon entropy across the 16S rRNA gene based on the alignment of representative sequences for each species present in the Greengenes database is shown in the upper part. A schematic representation of the variable regions of the 16S rRNA (gray) and commonly used primer-binding sites for amplicon sequencing (red) is depicted in the lower part. Adapted from (Johnson et al., 2019).

ATP. Sequencing is performed using sequential washing steps with one of the four nucleotides for incorporation by the DNA polymerase and measuring the incorporation of the nucleotide through light emission by luciferase using a charge-coupled device (CCD) camera. This method requires deoxyadenosine alfa-thio triphosphate instead of dATP since this nucleotide is incorporated by the DNA polymerase but not recognized by luciferase.

Solexa/Illumina sequencing uses fragmented template DNA and ligates these to oligonucleotide adapters, which are complementary to anchors on the flow-cell. Under limiting-dilution conditions, single-stranded template DNA is hybridized to the flow-cell, and bridge amplification is performed. Bridge amplification is a solid-phase PCR and has this name because adjacent anchor nucleotides lead to an arching of the single-stranded molecules that are then amplified to double-stranded “bridges” by PCR. This process is repeated until dense clusters of clonal populations of the initial flow-cell bound DNA strands are formed. The actual sequencing starts with the hybridization of a primer to the adapter sequence, after which the polymerase and a mixture of the four nucleotides tagged with different fluorophores and a reversible modification for chain termination are added. After a washing step, the fluorescence of the incorporated nucleotides is recorded with a CCD camera and subsequently, the fluorophores and the terminator are cleaved and washed away before the next sequencing cycle

is performed.

Ion Torrent uses a similar principle as the pyrosequencing approach. First, beads are populated with clonal DNA sequences using emulsion PCR and washed over a picowell plate. Then the four nucleotides are sequentially added and washed away for each sequencing cycle. The main difference comprises the detection technology for incorporated nucleotides. Ion Torrent uses ion semiconductor technology, which measures the difference in pH due to the release of protons during the polymerization of the nucleotides with a complementary metal-oxide-semiconductor (CMOS).

Amplicon sequencing of marker genes, such as the 16S rRNA gene, is often used to characterize microbial community composition. The 16S rRNA gene is suitable for this purpose because it is present in all bacterial genomes and contains both variable regions for the differentiation of different taxa and conserved regions as primer targets (**Figure 1.4**). However, it has also been shown that not all variable regions are equally suitable for distinguishing all taxonomic groups and that for a better classification on species and strain level, other marker genes or the full-length 16S rRNA gene may be more suitable (Lan et al., 2016; Yang et al., 2016; Johnson et al., 2019). Library preparation for amplicon sequencing can be achieved using PCR amplification of the marker gene and subsequent ligation of barcoded adapter sequences or by adding the adapter sequences and barcodes using fusion primers for PCR (Golebiewski and Tretyn, 2020). After sequencing, the raw NGS reads are prepared for downstream analysis by removing adaptor sequences, noise reduction, chimera removal, and quality filtering. The bioinformatics analysis of the clean reads includes clustering the reads to operational taxonomic units (OTUs) or amplicon sequence variants (ASVs), followed by taxonomic classification and subsequent statistical analysis. For a detailed discussion of different approaches and tools, current challenges in the analysis of amplicon sequencing data, see Chapter 4, or, e.g., the following references (Gloor et al., 2017; Boers et al., 2019; Leite and Kuramae, 2020; Llorens-Rico et al., 2021).

1.8 Methods for studying the microbiota of cheese

Classic microbiology methods, such as plating on selective media, have been the standard method for a long time to study the microbiota of cheese. These methods are still partly in use today, especially to detect pathogens for which specific culture media are available. For most other bacterial species, separation based on selective media is not feasible. Therefore, specific microbial groups were usually only roughly distinguished, for example, in categories such as aerobic mesophilic bacteria, lactococci, lactobacilli, enterococci, coliforms, and yeasts and molds. To identify the species, isolation of colonies on partially selective media and subsequent identification, e.g., by sequencing part of the 16S rRNA gene, is required. However, this procedure is relatively laborious, and species from subdominant populations are difficult to detect. The study of bacterial community composition and population dynamics in cheese has greatly improved with the advent of culture-independent molecular techniques. Methods such as denaturing gradient gel electrophoresis (DGGE), temporal temperature gradient gel electrophoresis (TTGE), single strand conformational polymorphism (SSCP), length heterogeneity PCR (LH-PCR), and terminal restriction fragment length polymorphism (T-RFLP) have been widely used in the past to study the microbial composition of cheese (Quigley et al., 2011; Cocolin et al., 2013). In the following section, a summary of the most commonly used culture-independent methods for the study of the cheese microbiome during the past decade is presented. The main advantages and limitations of these methods are listed in **Table 1.3**.

1.8.1 Quantitative real-time polymerase chain reaction

Initially, qPCR assays were primarily developed for the detection of pathogens, such as *Salmonella* spp. (Cheng et al., 2009), *Listeria monocytogenes* (Rossmanith et al., 2006; O'Grady et al., 2009), or *Staphylococcus aureus* (Alarcon et al., 2006) in dairy and other food products. In the last decade, numerous reverse transcription (RT-) qPCR and qPCR assays have been developed to quantify various bacteria used as starters and adjunct cultures or bacteria relevant for cheese ripening and food safety. For example, new qPCR assays were developed or optimized for *lacZ*-positive *Enterobacteriaceae* (Martin et al., 2010), *L. paracasei* (Falentin et al., 2010), *P. freudenreichii* (Falentin et al., 2010; Turgay et al., 2016), *Lc. lactis* (Ruggirello et al., 2014), *Acidipropionibacterium* (*Ap.*) *acidipropionici*, *Ap. jensenii*, *Ap. thoenii* (Turgay et al., 2016), *Staphylococcus aureus* genotype B (Sartori et al., 2017), *T. thermophilus* (Quigley et al., 2016), *L. helveticus* (Moser et al., 2017), and *L. parabuchneri* (Berthoud et al., 2017).

In different cheeses, the metabolically active fraction of bacteria during the ripening process was measured using RT-qPCR targeting a partial sequence of the 16S rRNA. Species monitored during ripening were, for example, *S. thermophilus* (Pega et al., 2016) and *Lc. lactis* (Ruggirello et al., 2016; Ruggirello et al., 2018), but also consortia of NSLAB comprising *L. paracasei*, *L. buchneri/parabuchneri*, *L. rhamnosus*, *L. brevis,* and

Loigolactobacillus coryniformis (Desfosses-Foucault et al., 2013). Parallel quantification of multiple species was also applied to natural whey starters and cheeses during the ripening. In natural whey starters for the manufacture of Grana Padano *S. thermophilus*, *L. helveticus*, *L. delbrueckii,* and *L. fermentum* were quantified during several back-slopping cycles (Bertani et al., 2020). Several probiotic strains and other LAB species were monitored during the ripening of cheddar cheese. For this, Ganesan et al. (2014) used species-specific qPCR assays for *L. acidophilus*, *L. casei/paracasei,* and *Bifidobacterium animalis* subsp. *lactis* and genus-specific assays for the *Lactobacillus* and *Lactococcus* genera. Xue et al. (2021) measured *Turibacter* spp., *L. helveticus,* and *L. fermentum* using qPCR to complement the NGS method with quantitative data to study slit defects in cheddar cheese. Specific primer pairs for the tyrosine decarboxylase gene (*tdc*) for *Enterococcus* spp., *L. brevis,* and *L. curvatus* were successfully used to quantify *tdc*-positive strains in commercial cheese samples (Ladero et al., 2010b). Further, the genera, *Advenella, Psychrobacter, Psychroflexus* (Schmitz-Esser et al., 2018), and *Brevibacterium* (Anast et al., 2019) were quantified using genus-specific qPCR assays during the ripening process in the rind of hard washed-rind cheese from Austria.

1.8.2 Denaturing gradient gel electrophoresis

DGGE allows the creation of profiles of unknown bacterial communities. Hence, a part of the 16S rRNA gene is amplified by PCR, and the PCR products are separated using electrophoresis in polyacrylamide gels with a gradient of denaturants (urea, formamide). The resolution is rather high, and many bacterial species/groups can be distinguished (Myers et al., 1987). Identification of the bacterial species can be achieved by comparison with known reference strains in the same gel or by excision of the bands and subsequent sequencing. Similarly, metabolically active species can also be detected and identified via 16S rRNA using a technique called reverse transcription PCR-DGGE (RT-PCR-DGGE). During the past decade, DGGE and RT-PCR-DGGE have often been used to complement culture-dependent identification methods to study the microbiota of the cheese core and the surface of smear-ripened cheeses. There are few studies in which the results of the culture-dependent methods were confirmed using DGGE (Fuka et al., 2010; Mei et al., 2014); on the other hand, several studies have found differences in the detected bacterial species (Quigley et al., 2011). For example, in a study of microbial diversity during the production and ripening of Castelmagno PDO cheese, *L. helveticus* was detected using DGGE but not by culturing on selective agar. Conversely, *L. plantarum* was exclusively identified on selective agar (Dolci et al., 2010). In more recent studies, lactobacilli were underrepresented in the DGGE profiles compared to plating techniques (Alegria et al., 2011; Aquilanti et al., 2011; Pangallo et al., 2014), while *S. thermophilus* was only detected using DGGE (Alegria et al., 2011; Gori et al., 2013). The DGGE method had advantages, especially for examining the microbial population on the surface of smear-ripened cheeses. Species, such as *Vagococcus carniphilus*, *Psychrobacter* spp., and *L. curvatus* (Gori et al., 2013) or *Lacticaseibacillus casei/paracasei*, *Staphylococcus equorum*, *Bacillus* sp., *Brevibacterium* sp., *Halomonas* sp., *Acinetobacter* sp., *Alkalibacterium* sp., and *Corynebacterium casei* (Yunita and Dodd, 2018) were additionally identified on the cheese surface using DGGE. Especially RT-PCR-DGGE targeting part of the 16S rRNA could provide a more detailed picture regarding the active microbiota on the cheese surface (Dolci et al., 2013; Dolci et al., 2014). TTGE, a similar method that uses a temperature gradient to separate DNA fragments, has also been successfully used to characterize bacterial communities of the surface of smear-ripened cheeses (Roth et al., 2010). A disadvantage of these methods targeting the 16S rRNA or the 16S rRNA gene is the limited resolution of closely related DNA sequences (Kisand and Wikner, 2003). The selection of the variable region of the 16S rRNA gene affects the ability to accurately identify species (Yunita and Dodd, 2018). Most of the abovementioned studies have used the V3 variable region of the 16S rRNA gene. Apart from community profiling using the 16S rRNA gene variable region, other target genes were used for DGGE. For example, the *slpH* gene that encodes an S-layer protein for the differentiation of *L. helveticus* strains (Miragoli et al., 2020) or the *hdcA* gene that encodes a histidine decarboxylase for the identification of histamine-producing bacteria in cheese were used (Diaz et al., 2016b).

1.8.3 Length-heterogeneity polymerase chain reaction

Length-heterogeneity polymerase chain reaction (LH-PCR) is a method that uses naturally occurring differences in the lengths of amplicons of the variable 16S rRNA gene region to identify microbial groups and species. The measured peaks must be compared with a database of known fragment lengths of reference species. LH-PCR can be partially automated using capillary electrophoresis and analysis software instead of a polyacrylamide gel. Garofalo et al. (2017) compared DGGE and LH-PCR as methods to study bacterial diversity in different foods, such as tofu, soy, butter, milk, cheese, salami, olives, and tomatoes. In this study, only a few LAB species, namely *L. delbrueckii* and *L. helveticus*, were detected using both methods. *S. thermophilus*, *Lc. lactis*, *E. faecalis*, *L. rhamnosus*, *P. acidilactici*, *L. fermentum,* and *L. plantarum* were detected exclusively using LH-PCR, while *L.*

crispatus and *L. sakei* were detected exclusively using DGGE. The authors concluded that LH-PCR had greater sensitivity and could detect many LAB species in dairy products, but that an expansion of the LH-PCR database would be desirable for analyzing other foods (Garofalo et al., 2017). Culture-dependent identification of species (RFLP) and LH-PCR were used to study the diversity of lactic acid bacteria in hard cheese (Grana Padano) made from raw milk (Pogacic et al., 2013). *S. thermophilus*, *L. delbrueckii*, *L. helveticus*, *L. fermentum*, *L. rhamnosus*, *L. casei*, *L. paracasei*, *Lc. lactis*, and *P. acidilactici* were identified by both approaches, while *E. faecalis* and *L. paraplantarum* could only be detected using the culture-dependent approach. A study of artisanal Minas cheese from Brazil using culture-dependent and independent methods identified 20 different species in all cheeses studied. From these 20 species *Lc. lactis* subsp. *lactis*, *L. curvatus*, *L. fermentum*, *L. delbrueckii*, *S. thermophilus*, *P. acidilactici*, and *L. mesenteroides* were exclusively detected using LH-PCR (Perin et al., 2017). All studies using LH-PCR mentioned above have observed additional unidentified peaks. A possible explanation is that the database did not contain all species. However, Savo Sardaro et al. (2018) reported different profiles for different strains when the annealing temperatures of the primers were too low. They further reported highly similar peaks for *Enterococcus* spp. and *L. delbrueckii* (329–330 bp), *Weissella* spp. and *P. pentosaceus* (346 bp) using an optimized annealing temperature for PCR. Similarly, a study focusing on the lysis of starter LAB and NSLAB also reported difficulties in species differentiation. They used LH-PCR targeting the V1–V2 region of the 16S rRNA gene, the differentiation of the species of the *L. rhamnosus* group (*L. casei/paracasei* and *L. rhamnosus*) and the differentiation between *S. thermophilus* and *Lc. lactis* was not possible (Santarelli et al., 2013).

Table 1.3: Advantages and limitations of methods used to study cheese microbiology

Method	Level	Advantages	Limitations
Plating techniques	Bacterial groups (e.g., aerobic / anaerobic, salt tolerant, carbon sources, and antibiotics resistance)	Viable (cultivable)/dead discrimination and isolates	Cultivation dependent, slow, elaborate, limited selectivity of nutrient media, and limited identification of sub-populations
qPCR	Bacterial groups / genes (targeted)	Cultivation-independent, quantitative data, fast, and simple data interpretation	Elaborate primer design and targeted approach
DGGE	Bacterial composition	Cultivation-independent, identification of unknown bands by sequencing	Reference DNA, elaborate, not automated, limited resolution (partial 16S rRNA gene), and no quantitative data
LH-PCR	Bacterial composition	Cultivation-independent, partial automation (capillary electrophoresis and analysis software)	Reference DNA, elaborate, limited resolution (partial 16S rRNA gene), and no quantitative data
NGS (16S rRNA gene)	Bacterial composition	Cultivation-independent, high-throughput, automated, untargeted approach, potential identification of unknown taxa, and semi-quantitative data	Depends on reference database quality, sophisticated analysis, limited species resolution, and compositional data

1.8.4 Next-generation sequencing

Microbiome analyses in fermented foods have become much more detailed with the advent and proliferation of new high-throughput sequencing technologies. In particular, 16S rRNA gene amplicon sequencing has been used frequently for cheese analysis recently. Compared to the previously discussed culture-independent methods, next-generation sequencing (NGS) has several advantages. The high-throughput allows obtaining a large amount of

information in a short time. Further, the read counts for different taxonomic groups estimate the relative abundances in the measured sample.

In a broad study of Irish artisanal cheese, examining cheese rinds and cores made with pasteurized or unpasteurized milk demonstrated the potential of 16S rRNA gene amplicon sequencing. In this study, the cheese microbiome was investigated in great detail and revealed genera that had never been detected in cheese before (Quigley et al., 2012). In a study of Polish Oscypek cheese with culture-dependent methods, DGGE, and 454 pyrosequencing, all three methods agreed that the predominant genera were *Lactococcus*, *Lactobacillus*, *Leuconostoc,* and *Streptococcus* (Alegria et al., 2012). *E. durans*, *E. italicus,* and *L. parabuchneri* were identified solely using the culture-dependent method. *Streptococcus* spp., other than *S. thermophilus, Lc. Raffinolactis,* and *L. helveticus* were only identified using DGGE. The pyrosequencing approach revealed additional bacteria belonging to the *Bifidobacteriaceae* and *Moraxellaceae* families. However, there were also many unclassified reads (~20%), and taxonomic classification of the reads could only be performed to the genus level due to technological limitations, such as short amplicon lengths and incomplete databases. RT-PCR-DGGE and pyrosequencing of the V1–V3 16S rRNA variable regions were used to study the microbiota in Fontina PDO cheese (Dolci et al., 2014). Thereby, the dominant starter bacteria, *S. thermophilus* and *L. delbrueckii*, were detected using both methods. However, there were some differences between the results of various methods for some NSLAB species. For example, *L. plantarum* and *E. faecium* were detected only using RT-PCR-DGGE, while *E. faecalis* was only detected using pyrosequencing. Several studies used (RT-PCR-)DGGE and NGS as complementary methods for species identification and to investigate the relative abundances of the major and minor bacterial groups, respectively (Ryssel et al., 2015; Alessandria et al., 2016; Cardinali et al., 2017; Ramezani et al., 2017). Improvements in sequencing techniques, optimized protocols, and databases have led to an increasing number of species-level assignments with 16S rRNA gene amplicon sequencing. For example, Bassi et al. (2015) assigned 94.3% of all reads to a species using an Illumina NGS workflow to analyze the microbiota in Italian cheeses with late blowing defects.

High-throughput 16S rRNA gene amplicon sequencing has evolved rapidly; it has been used in countless studies and is now considered the gold standard for studying microbial population dynamics and bacterial communities. Amplicon sequencing has been used to investigate microbial contaminants in raw milk and the proliferation and survival of starter and contaminant bacteria in a Danish cheese (Masoud et al., 2012) or to reveal differences in the composition of the microbiota in the cheese core of five regional Grana Padano production regions (Zago et al., 2021). Amplicon sequencing provided new insights into the composition of the cheese rind microbiome. For example, the use of this approach has shown that the house microbiome of cheese plants has a major influence on the rind microbiota of washed-rind cheeses (Bokulich and Mills, 2013) or that microbial diversity correlates strongly with rind type and moisture but only weakly with milk treatment or milk origin (Wolfe et al., 2014). Further, amplicon sequencing has been used to study the microbial composition and dynamics in starter cultures, e.g., in undefined milk starters from raw milk for pasta-filata cheeses (Parente et al., 2016) or natural whey starters for Parmigiano Reggiano cheese (Bertani et al., 2020). Microbial dynamics in cheese during ripening have also been intensively investigated. The studies investigated, for example, the influence of NSLAB during the ripening of Grana Padano cheese (Levante et al., 2017), the influence of autochthonous bacteria as starters in Pecorino Crotonese cheese (De Pasquale et al., 2019), the development of the cheese core and rind microbiota in a continental-type cheese (O'Sullivan et al., 2015a) or the effect of milk centrifugation on *Clostridium* spp. Spores and the microbial composition during ripening in a Maasdam cheese (Lamichhane et al., 2018). The investigation and identification of causes for more complex cheese defects caused or intensified by a consortium of different species have been facilitated using amplicon sequencing. For example, the identification of *T. thermophilus* as the cause of the pink discoloration defect in cheddar cheese and the starter composition as an enhancer of the discoloration (Quigley et al., 2016), or the identification of bacterial consortia associated with slit defects in cheddar cheese (Xue et al., 2021).

The 16S rRNA gene is the most targeted gene for studying microbial populations in cheese; however, other target genes have also been used to study specific sub-populations. A high-throughput amplicon sequencing approach targeting the histidine and tyrosine decarboxylase genes was used to study the biogenic amine-producing bacteria in several raw milk cheese varieties (O'Sullivan et al., 2015b). Amplicon sequencing targeting the pyruvate oxidase gene (*spxB*) increased the separation of the species of the *L. casei* group in a study of the NSLAB population in Grana Padano cheeses (Levante et al., 2017).

In addition to the usage of amplicon sequencing, shotgun metagenomics approaches are increasingly applied in microbiology. With this technique, whole genomes of microbes can be directly determined from the sample by shotgun sequencing and metagenome assembly. The bioinformatic analysis is much more complex and data-intensive than for amplicon sequencing, but genetic variations and the frequency of functional genes of the sequenced microbes can be studied in detail (Escobar-Zepeda et al., 2015).

1.8.5 Bioinformatic approaches to overcoming the limitations of current NGS methods

High-throughput amplicon sequencing is a well-established approach for investigating microbial compositions, but it still has limitations. One of the main concerns is the semi-quantitative and compositional nature of the data resulting from the limited number of reads for a run in current sequencing machines (Gloor et al., 2017). Due to this, observed dynamics of relative abundances and the actual change in numbers of taxa in a sample may differ (Props et al., 2017). This can also make it difficult to compare data from different studies using different sequencers. There are several bioinformatic approaches for solving such problems, e.g., count normalization or rarefaction for the standardization and comparison of samples. However, these approaches also have drawbacks, such as a loss of information, and it remains difficult to compare communities with significant differences in cell density, even if equimolar pooling of DNA ensures the same size of sequencing libraries in all experiments.

Compositional approaches use log transformations, such as the centered log-ratio transformation (Aitchison, 1982), to calculate the geometric mean for multivariate tests; the zero values in sparse sequencing data have to be replaced, deleted, or estimated (Martino et al., 2019). Nowadays, there are various software tools, and various models are available for compositional data analysis (Quinn et al., 2019; Leite and Kuramae, 2020). If the experimental approach is optimized (e.g., sampling, DNA extraction, and normalization), these tools, with the right assumptions and an appropriate model for the specific scientific question, can avoid or improve some systematic errors and spurious correlations. A recent study benchmarked the performance of several commonly used analytical approaches based on simulated datasets with varying sampling depths (Llorens-Rico et al., 2021). The authors of the study simulated three ecological scenarios (Succession, Blooming and Dysbiosis) with samples of different microbial densities. Their results suggest that experimental quantitative cell density data (e.g., by spike-ins, qPCR, or flow cytometry) can be used to bypass compositionality and that these approaches outperform computational approaches to identify taxon-metadata associations or to study taxon-taxon interactions, especially when taxa distributions become increasingly heterogeneous between samples.

Advancements in sequencing and analytical techniques are likely soon to overcome the limitations of the current NGS methods and will provide deeper insights into microbial community dynamics; however, bioinformatic analysis will also become increasingly important for the exploitation of larger and more complex data sets.

1.8.6 Current and future developments in food microbiology

A third generation of sequencing technologies is nowadays available and brings many advantages, such as longer reads, real-time sequencing, amplification-free, and single-molecule sequencing. This results in a massive gain for both metabarcoding and metagenomics (whole-genome shotgun) approaches (Bharti and Grimm, 2021). A combination of NGS and a third-generation sequencing metagenomics approach in natural whey cultures has shown that this approach results in finished genomes and provides a strain-level resolution for the dominant bacterial species (Somerville et al., 2019). Such data allow a more detailed study of bacterial communities, revealing information about functional genes, metabolic networks, and taxa-taxa interactions. Especially the combination of different omics approaches offers many opportunities for functional profiling and in-depth biological insights. The potential for an improved understanding of cheese quality and flavor by multi-omics approaches was recently reviewed (Afshari et al., 2018).

A recent example is the combination of 16S rRNA amplicon sequencing data with metabolome data from gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS) in artisanal and industrial cheddar cheeses (Afshari et al., 2020). The authors reported numerous correlations between specific genera and metabolites to clearly distinguish between artisanal and industrially produced cheese based on the composition of the microbiota and the metabolite profiles. Such studies could provide interesting information for the selection of strains for starter and adjunct cultures. Yang et al. (2021) used a deep metagenome sequencing approach and LC/MS to analyze Chinese and European cheeses. The strain-level resolution revealed a great variation in the strains of the same species, and the authors suggested that cheese flavor and quality might be determined by specific strains rather than species-level composition. Identifying correlations between metabolites and specific strains will be highly useful for future research in cheese microbiology and the development of starter and adjunct cultures.

1.9 Investigation of cheese quality defects

For some defects in cheese, the cause is clearly identified, for example, the cause of late blowing by *C. tyrobutyricum*. While for others, various possible sources have been discussed in the literature, such as the involvement of coliform bacteria, yeasts, or NSLAB populations in early gas formation. At Agroscope (the Swiss

Confederation's centre of excellence for agricultural research), cheese defects are investigated mainly by measuring the chemical and biochemical parameters of the cheeses, including water, fat, salt and mineral content, volatile carboxylic acids, protein and amino acid profiles, biogenic amine content, pH, lactose, lactate, citrate concentration, and enzymatic activity of specific marker lipases and peptidases. Classic microbial analyses were used to detect groups of microbes on a selective agar medium. The microbial groups detected were, for example, total aerobic mesophilic cell count, psychrotrophs, salt-tolerant bacteria, enterococci, bacteria with proteolytic or lipolytic activity, and facultative heterofermentative lactobacilli. Further, pathogens such as *Listeria monocytogenes*, *Salmonella* spp., *Staphylococcus aureus,* and *Escherichia coli* O157 were detected. Therefore, many defects are well characterized with symptoms and influence on the chemical and biochemical composition but poorly connected to the microbial cause on the species level. For example, in cheeses with gas defects, a volatile carboxylic acid profile is made to indicate if butyric acid (*C. tyrobutyricum*), propionic acid (PAB), or formic acid (FHL / coliforms) levels are elevated. This metabolic approach was recently complemented by qPCR assays specific for *C. tyrobutyricum* (unpublished), PAB (Turgay et al., 2016), and *L. parabuchneri* (Berthoud et al., 2017). Often defects are not caused by one species or strain; examples include the interaction of *Paucilactobacillus wasatchensis* and *S. thermophilus* in gas formation or the pink discoloration defect (Ortakci et al., 2015; Quigley et al., 2016), where the composition of the starter culture influences the intensity of the discoloration.

The simultaneous monitoring of various species offers new possibilities for investigating microbial communities in cheeses and could contribute to the faster recognition of cheese defects. The indirect approach of detecting metabolites of species involved in cheese quality defects could be extended with the detection of the microbial species. The detection at the species level can indicate potential problems before the accumulation of detrimental metabolites. By monitoring different species simultaneously, synergistic or antagonistic trends can be evaluated, and their impact on cheese quality can be investigated in detail.

1.10 Objectives of the thesis

Molecular biology methods have largely replaced traditional culture-dependent methods in microbiology for the study of bacterial communities because they provide faster results, require less labor, and perform better for the detection of sub-populations of species in complex systems. The molecular biology methods used to characterize the bacterial composition of the cheese microbiota so far had the disadvantage that universal primers had to be found, which made it difficult to resolve highly similar bacterial groups. Neither previous methods, such as DGGE or LH-PCR, have generated quantitative compositional information due to technical constraints. Next-generation sequencing methods provide semi-quantitative and compositional data, and additional experimental quantitative cell density data can significantly improve comparisons between samples with different sampling depths. However, in practice, for cheese research and monitoring cheese quality, it is often important to know the composition and the cell density of the species detected, as the quantity of a species present can be indicative of the quality or severity of the defect. qPCR has long been used to detect and enumerate pathogens and microbiological causative agents of cheese defects. However, the development of qPCR assays requires considerable initial effort to design suitable primers, and a relatively low throughput limits the performance. Therefore, usually, only the most important bacteria for food safety and cheese quality were targeted individually. The development of microfluidic chips that significantly reduce the amount of material and work required for qPCR has made it possible to rapidly detect several bacterial targets in multiple samples in parallel. The bacterial community in cheese is shaped by harsh conditions during production and ripening, such as the heating of the milk at the beginning of production, the low pH due to rapid acidification by the starter cultures, and the salt content. This core community mainly comprises LAB; this limited diversity makes studying the complex microbial composition in cheese by high-throughput qPCR very appealing.

This thesis was part of a vision for the future assessment of cheese quality, namely, the use of recent molecular biology methods for an extensive and fast analysis of the cheese microbiota. The basic idea was to use next-generation sequencing to characterize the cheese microbiota and then HT-qPCR as a novel method for the fast quantification of the bacterial species relevant for cheese quality. Data from next-generation sequencing and scientific literature should provide information as to the most relevant target species. The overall goal was then to develop a HT-qPCR system for these species that could eventually be used as a fast, cost-efficient, and modular diagnostic tool to support cheese quality consultants, detect microbiological-related cheese defects, and help cheese producers monitor the microbial composition of their cheeses.

The broad objectives of this thesis were 1) to identify and develop the basis for a high-throughput qPCR system that would detect and quantify bacterial species relevant to cheese quality, 2) to develop and validate the

high-throughput qPCR system, and 3) to apply the developed system for the analysis of the microbial composition of cheeses.

More specifically, the first sub-objective was to identify and select bacterial species that influence cheese quality. These included the bacteria used in starter and adjunct cultures, known species responsible for fermentation defects or biogenic amine production, and NSLAB, which may contribute to flavor development or be involved in cheese defects. The next goal was to identify the HT-qPCR requirements, compile, and validate the appropriate primers for the detection and quantification of the intended bacteria and validate them on the selected HT-qPCR chip from Fluidigm. Thereafter, the application of HT-qPCR in real cheese samples was the basis to further evaluate costs, speed, accuracy, and sensitivity of the diagnostic approach.

1.11 References

Abeijón Mukdsi, M.C., Falentin, H., Maillard, M.-B., Chuat, V., Medina, R.B., Parayre, S., et al. (2014). The secreted esterase of *Propionibacterium freudenreichii* has a major role in cheese lipolysis. *Appl. Environ. Microbiol.* 80**,** 751-756. doi: 10.1128/AEM.03640-13.

Addis, E., Fleet, G.H., Cox, J.M., Kolak, D., and Leung, T. (2001). The growth, properties and interactions of yeasts and bacteria associated with the maturation of Camembert and blue-veined cheeses. *Int. J. Food Microbiol.* 69**,** 25-36.

Afshari, R., Pillidge, C.J., Dias, D.A., Osborn, A.M., and Gill, H. (2018). Cheesomics: the future pathway to understanding cheese flavour and quality. *Crit. Rev. Food Sci. Nutr.* 60**,** 33-47. doi: 10.1080/10408398.2018.1512471.

Afshari, R., Pillidge, C.J., Read, E., Rochfort, S., Dias, D.A., Osborn, A.M., et al. (2020). New insights into cheddar cheese microbiota-metabolome relationships revealed by integrative analysis of multi-omics data. *Sci. Rep.* 10**,** 3164. doi: 10.1038/s41598-020-59617-9.

Agroscope (2017). *Kulturen* [Online]. Available: https://www.agroscope.admin.ch/agroscope/de/home/themen/lebensmittel/qualitaet/kulturen/die-kulturen.html [Accessed 28.02.2017 2017].

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)* 44**,** 139-177.

Alarcon, B., Vicedo, B., and Aznar, R. (2006). PCR-based procedures for detection and quantification of *Staphylococcus aureus* and their application in food. *J. Appl. Microbiol.* 100**,** 352-364. doi: 10.1111/j.1365-2672.2005.02768.x.

Alegria, A., Gonzalez, R., Diaz, M., and Mayo, B. (2011). Assessment of microbial populations dynamics in a blue cheese by culturing and denaturing gradient gel electrophoresis. *Curr. Microbiol.* 62**,** 888-893. doi: 10.1007/s00284-010-9799-7.

Alegria, A., Szczesny, P., Mayo, B., Bardowski, J., and Kowalczyk, M. (2012). Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches. *Appl. Environ. Microbiol.* 78**,** 1890-1898. doi: 10.1128/AEM.06081-11.

Alessandria, V., Ferrocino, I., De Filippis, F., Fontana, M., Rantsiou, K., Ercolini, D., et al. (2016). Microbiota of an Italian Grana-like cheese during manufacture and ripening, unraveled by 16S rRNA-based approaches. *Appl. Environ. Microbiol.* 82**,** 3988-3995. doi: 10.1128/AEM.00999-16.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215**,** 403-410. doi: 10.1016/s0022-2836(05)80360-2.

Anast, J.M., Dzieciol, M., Schultz, D.L., Wagner, M., Mann, E., and Schmitz-Esser, S. (2019). *Brevibacterium* from Austrian hard cheese harbor a putative histamine catabolism pathway and a plasmid for adaptation to the cheese environment. *Sci. Rep.* 9**,** 6164. doi: 10.1038/s41598-019-42525-y.

Aquilanti, L., Babini, V., Santarelli, S., Osimani, A., Petruzzelli, A., and Clementi, F. (2011). Bacterial dynamics in a raw cow's milk *Caciotta* cheese manufactured with aqueous extract of *Cynara cardunculus* dried flowers. *Lett. Appl. Microbiol.* 52**,** 651-659. doi: 10.1111/j.1472-765X.2011.03053.x.

Arques, J.L., Rodriguez, E., Langa, S., Landete, J.M., and Medina, M. (2015). Antimicrobial activity of lactic acid bacteria in dairy products and gut: effect on pathogens. *Biomed Res Int* 2015**,** 584183. doi: 10.1155/2015/584183.

Ascone, P., Maurer, J., Haldemann, J., Irmler, S., Berthoud, H., Portmann, R., et al. (2017). Prevalence and diversity of histamine-forming *Lactobacillus parabuchneri* strains in raw milk and cheese A case study. *Int. Dairy J.* 70**,** 26-33. doi: 10.1016/j.idairyj.2016.11.012.

Bassi, D., Puglisi, E., and Cocconcelli, P.S. (2015). Understanding the bacterial communities of hard cheese with blowing defect. *Food Microbiol.* 52**,** 106-118. doi: 10.1016/j.fm.2015.07.004.

Beresford, T., and Williams, A. (2004). "The Microbiology of Cheese Ripening," in *Cheese: Chemistry, Physics and Microbiology,* eds. P.F. Fox, P.L.H. McSweeney, T.M. Cogan & T.P. Guinee. Academic Press**,** 287-317.

Bernardeau, M., Guguen, M., and Vernoux, J.P. (2006). Beneficial lactobacilli in food and feed: long-term use, biodiversity and proposals for specific and realistic safety assessments. *FEMS Microbiol. Rev.* 30**,** 487-513. doi: 10.1111/j.1574-6976.2006.00020.x.

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

Berthoud, H., Wüthrich, D., Bruggmann, R., Wechsler, D., Fröhlich-Wyder, M.-T., and Irmler, S. (2017). Development of new methods for the quantitative detection and typing of *Lactobacillus parabuchneri* in dairy products. *Int. Dairy J.* 70**,** 65-71. doi: 10.1016/j.idairyj.2016.10.005.

Bharti, R., and Grimm, D.G. (2021). Current challenges and best-practice protocols for microbiome analysis. *Brief. Bioinform.* 22**,** 178-193. doi: 10.1093/bib/bbz155.

Boers, S.A., Jansen, R., and Hays, J.P. (2019). 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* 38**,** 1059-1070. doi: 10.1007/s10096-019-03520-3.

Bokulich, N.A., and Mills, D.A. (2013). Facility-specific "house" microbiome drives microbial landscapes of artisan cheesemaking plants. *Appl. Environ. Microbiol.* 79**,** 5214-5223. doi: 10.1128/AEM.00934-13.

Breslauer, K.J., Frank, R., Blöcker, H., and Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U. S. A.* 83**,** 3746-3750. doi: 10.1073/pnas.83.11.3746.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55**,** 611-622. doi: 10.1373/clinchem.2008.112797.

Cardinali, F., Osimani, A., Taccari, M., Milanovic, V., Garofalo, C., Clementi, F., et al. (2017). 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.* 255**,** 7-16. doi: 10.1016/j.ijfoodmicro.2017.05.018.

Cheng, C.-M., Van, K.T., Lin, W., and Ruby, R.M. (2009). Interlaboratory validation of a real-time PCR 24-hour rapid method for detection of *Salmonella* in foods. *J. Food Prot.* 72**,** 945-951. doi: 10.4315/0362-028x-72.5.945.

Chuang, L.Y., Cheng, Y.H., and Yang, C.H. (2013). Specific primer design for the polymerase chain reaction. *Biotechnol. Lett.* 35**,** 1541-1549. doi: 10.1007/s10529-013-1249-8.

Claesson, M.J., van Sinderen, D., and O'Toole, P.W. (2008). *Lactobacillus* phylogenomics towards a reclassification of the genus. *Int. J. Syst. Evol. Microbiol.* 58**,** 2945-2954. doi: 10.1099/ijs.0.65848-0.

Cocolin, L., Alessandria, V., Dolci, P., Gorra, R., and Rantsiou, K. (2013). Culture independent methods to assess the diversity and dynamics of microbiota during food fermentation. *Int. J. Food Microbiol.* 167**,** 29-43. doi: 10.1016/j.ijfoodmicro.2013.05.008.

Cogan, T.M., Goerges, S., Gelsomino, R., Larpin, S., Hohenegger, M., Bora, N., et al. (2014). Biodiversity of the Surface Microbial Consortia from Limburger, Reblochon, Livarot, Tilsit, and Gubbeen Cheeses. *Microbiol. Spectrum* 2**,** 2.1.22. doi: doi:10.1128/microbiolspec.CM-0010-2012.

Crow, V.L. (1986). Utilization of lactate isomers by *Propionibacterium freudenreichii* subsp. *shermanii*: Regulatory role for intracellular pyruvate. *Appl. Environ. Microbiol.* 52**,** 352-358. doi: 10.1128/AEM.52.2.352-358.1986.

de las Rivas, B., Marcobal, Á., Carrascosa, A.V., and Muñoz, R. (2006). PCR detection of foodborne bacteria producing the biogenic amines histamine, tyramine, putrescine, and cadaverine. *J. Food Prot.* 69**,** 2509-2514. doi: 10.4315/0362-028X-69.10.2509.

de Oliveira, G.B., Favarin, L., Luchese, R.H., and McIntosh, D. (2015). Psychrotrophic bacteria in milk: How much do we really know? *Braz. J. Microbiol.* 46**,** 313-321. doi: 10.1590/S1517-838246220130963.

De Pasquale, I., Di Cagno, R., Buchin, S., De Angelis, M., and Gobbetti, M. (2019). Use of autochthonous mesophilic lactic acid bacteria as starter cultures for making Pecorino Crotonese cheese: Effect on compositional, microbiological and biochemical attributes. *Food Res. Int.* 116**,** 1344-1356. doi: 10.1016/j.foodres.2018.10.024.

Deeth, H.C. (2006). Lipoprotein lipase and lipolysis in milk. *Int. Dairy J.* 16**,** 555-562. doi: 10.1016/j.idairyj.2005.08.011.

del Rio, B., Ladero, V., Redruello, B., Linares, D.M., Fernandez, M., Martin, M.C., et al. (2015). Lactose-mediated carbon catabolite repression of putrescine production in dairy *Lactococcus lactis* is strain dependent. *Food Microbiol.* 48**,** 163-170. doi: 10.1016/j.fm.2014.11.018.

Desfosses-Foucault, E., LaPointe, G., and Roy, D. (2013). Dynamics and rRNA transcriptional activity of lactococci and lactobacilli during Cheddar cheese ripening. *Int. J. Food Microbiol.* 166**,** 117-124. doi: 10.1016/j.ijfoodmicro.2013.06.022.

Diaz, M., Ladero, V., Del Rio, B., Redruello, B., Fernandez, M., Martin, M.C., et al. (2016a). Biofilm-forming capacity in biogenic amine-producing bacteria isolated from dairy products. *Front. Microbiol.* 7**,** 591. doi: 10.3389/fmicb.2016.00591.

Diaz, M., Ladero, V., Redruello, B., Sanchez-Llana, E., del Rio, B., Fernandez, M., et al. (2016b). A PCR-DGGE method for the identification of histamine-producing bacteria in cheese. *Food Control* 63**,** 216-223. doi: 10.1016/j.foodcont.2015.11.035.

Dieffenbach, C.W., Lowe, T.M., and Dveksler, G.S. (1993). General concepts for PCR primer design. *PCR Methods Appl.* 3**,** S30-37. doi: 10.1101/gr.3.3.s30.

Dolci, P., Alessandria, V., Rantsiou, K., Bertolino, M., and Cocolin, L. (2010). Microbial diversity, dynamics and activity throughout manufacturing and ripening of Castelmagno PDO cheese. *Int. J. Food Microbiol.* 143**,** 71-75. doi: 10.1016/j.ijfoodmicro.2010.07.007.

Dolci, P., De Filippis, F., La Storia, A., Ercolini, D., and Cocolin, L. (2014). rRNA-based monitoring of the microbiota involved in Fontina PDO cheese production in relation to different stages of cow lactation. *Int. J. Food Microbiol.* 185**,** 127-135. doi: 10.1016/j.ijfoodmicro.2014.05.021.

Dolci, P., Zenato, S., Pramotton, R., Barmaz, A., Alessandria, V., Rantsiou, K., et al. (2013). Cheese surface microbiota complexity: RT-PCR-DGGE, a tool for a detailed picture? *Int. J. Food Microbiol.* 162**,** 8-12. doi: 10.1016/j.ijfoodmicro.2012.12.009.

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

Escobar-Zepeda, A., Vera-Ponce de Leon, A., and Sanchez-Flores, A. (2015). The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front Genet* 6**,** 348. doi: 10.3389/fgene.2015.00348.

Falentin, H., Henaff, N., Le Bivic, P., Deutsch, S.M., Parayre, S., Richoux, R., et al. (2012). Reverse transcription quantitative PCR revealed persistency of thermophilic lactic acid bacteria metabolic activity until the end of the ripening of Emmental cheese. *Food Microbiol.* 29**,** 132-140. doi: 10.1016/j.fm.2011.09.009.

Falentin, H., Postollec, F., Parayre, S., Henaff, N., Le Bivic, P., Richoux, R., et al. (2010). Specific metabolic activity of ripening bacteria quantified by real-time reverse transcription PCR throughout Emmental cheese manufacture. *Int. J. Food Microbiol.* 144**,** 10-19. doi: 10.1016/j.ijfoodmicro.2010.06.003.

Fox, P.F., Guinee, T.P., Cogan, T.M., and McSweeney, P.L.H. (2017a). *Fundamentals of Cheese Science.* Springer US.

Fox, P.F., Guinee, T.P., Cogan, T.M., and McSweeney, P.L.H. (2017b). "Microbiology of Cheese Ripening," in *Fundamentals of Cheese Science*. Springer, Boston, MA**,** 333-390.

Fox, P.F., Guinee, T.P., Cogan, T.M., and McSweeney, P.L.H. (2017c). "Starter Cultures," in *Fundamentals of Cheese Science*. Springer, Boston, MA).

Fröhlich-Wyder, M.-T., Bisig, W., Guggisberg, D., Irmler, S., Jakob, E., and Wechsler, D. (2015). Influence of low pH on the metabolic activity of *Lactobacillus buchneri* and *Lactobacillus parabuchneri* strains in Tilsit-type model cheese. *Dairy Science & Technology* 95**,** 569-585. doi: 10.1007/s13594-015-0238-1.

Fröhlich-Wyder, M.-T., Guggisberg, D., Badertscher, R., Wechsler, D., Wittwer, A., and Irmler, S. (2013). The effect of *Lactobacillus buchneri* and *Lactobacillus parabuchneri* on the eye formation of semi-hard cheese. *Int. Dairy J.* 33**,** 120-128. doi: 10.1016/j.idairyj.2013.03.004.

Fuka, M.M., Engel, M., Skelin, A., Redzepovic, S., and Schloter, M. (2010). Bacterial communities associated with the production of artisanal Istrian cheese. *Int. J. Food Microbiol.* 142**,** 19-24. doi: 10.1016/j.ijfoodmicro.2010.05.008.

Ganesan, B., Stuart, M.R., and Weimer, B.C. (2007). Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 73**,** 2498-2512. doi: 10.1128/AEM.01832-06.

Ganesan, B., and Weimer, B.C. (2017). "Amino acid catabolism and its relationship to cheese flavor outcomes," in *Cheese***,** 483-516.

Ganesan, B., Weimer, B.C., Pinzon, J., Dao Kong, N., Rompato, G., Brothersen, C., et al. (2014). Probiotic bacteria survive in Cheddar cheese and modify populations of other lactic acid bacteria. *J. Appl. Microbiol.* 116**,** 1642-1656. doi: 10.1111/jam.12482.

Gardini, F., Tofalo, R., Belletti, N., Iucci, L., Suzzi, G., Torriani, S., et al. (2006). Characterization of yeasts involved in the ripening of Pecorino Crotonese cheese. *Food Microbiol.* 23**,** 641-648. doi: 10.1016/j.fm.2005.12.005.

Garofalo, C., Bancalari, E., Milanovic, V., Cardinali, F., Osimani, A., Sardaro, M.L.S., et al. (2017). Study of the bacterial diversity of foods: PCR-DGGE versus LH-PCR. *Int. J. Food Microbiol.* 242**,** 24-36. doi: 10.1016/j.ijfoodmicro.2016.11.008.

Gatti, M., Bottari, B., Lazzi, C., Neviani, E., and Mucchetti, G. (2014). Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. *J. Dairy Sci.* 97**,** 573-591. doi: 10.3168/jds.2013-7187.

Gatti, M., Fornasari, M.E., Mucchetti, G., Addeo, F., and Neviani, E. (1999). Presence of peptidase activities in different varieties of cheese. *Lett. Appl. Microbiol.* 28, 368-372. doi: 10.1046/j.1365-2672.1999.00541.x.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8**,** 2224. doi: 10.3389/fmicb.2017.02224.

Golebiewski, M., and Tretyn, A. (2020). Generating amplicon reads for microbial community assessment with next-generation sequencing. *J. Appl. Microbiol.* 128**,** 330-354. doi: 10.1111/jam.14380.

Gori, K., Ryssel, M., Arneborg, N., and Jespersen, L. (2013). Isolation and identification of the microbiota of Danish farmhouse and industrially produced surface-ripened cheeses. *Microb. Ecol.* 65**,** 602-615. doi: 10.1007/s00248-012-0138-3.

Green, M.R., and Sambrook, J. (2019). Polymerase Chain Reaction. *Cold Spring Harb Protoc* 2019. doi: 10.1101/pdb.top095109.

Gupta, A.K., and Gupta, U.D. (2020). "Chapter 20 - Next generation sequencing and its applications," in *Animal Biotechnology* (Second Edition), eds. A.S. Verma & A. Singh. (Boston: Academic Press)**,** 395-421.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107**,** 1-8. doi: 10.1016/j.ygeno.2015.11.003.

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 10**,** 5029. doi: 10.1038/s41467-019-13036-1.

Karametsi, K., Kokkinidou, S., Ronningen, I., and Peterson, D.G. (2014). Identification of bitter peptides in aged cheddar cheese. *J. Agric. Food Chem.* 62**,** 8034-8041. doi: 10.1021/jf5020654.

Kisand, V., and Wikner, J. (2003). Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *J. Microbiol. Methods* 54**,** 183-191. doi: 10.1016/s0167-7012(03)00038-1.

Klijn, N., Nieuwenhof, F.F., Hoolwerf, J.D., van der Waals, C.B., and Weerkamp, A.H. (1995). Identification of *Clostridium tyrobutyricum* as the causative agent of late blowing in cheese by species-specific PCR amplification. *Appl. Environ. Microbiol.* 61**,** 2919-2924.

Ladero, V., Calles-Enriquez, M., Fernandez, M., and Alvarez, M.A. (2010a). Toxicological effects of dietary biogenic amines. *Current Nutrition & Food Science* 6**,** 145-156. doi: 10.2174/157340110791233256.

Ladero, V., Fernandez, M., Cuesta, I., and Alvarez, M.A. (2010b). Quantitative detection and identification of tyramine-producing enterococci and lactobacilli in cheese by multiplex qPCR. *Food Microbiol.* 27**,** 933-939. doi: 10.1016/j.fm.2010.05.026.

Lamichhane, P., Pietrzyk, A., Feehily, C., Cotter, P.D., Mannion, D.T., Kilcawley, K.N., et al. (2018). Effect of milk centrifugation and incorporation of high heat-treated centrifugate on the microbial composition and levels of volatile organic compounds of Maasdam cheese. *J. Dairy Sci.* 101**,** 5738-5750. doi: 10.3168/jds.2017-14180.

Lan, Y., Rosen, G., and Hershberg, R. (2016). Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* 4**,** 18. doi: 10.1186/s40168-016-0162-5.

Larpin, S., Mondoloni, C., Goerges, S., Vernoux, J.-P., Guéguen, M., and Desmasures, N. (2006). *Geotrichum candidum* dominates in yeast population dynamics in Livarot, a French red-smear cheese. *FEMS Yeast Res.* 6**,** 1243-1253. doi: 10.1111/j.1567-1364.2006.00127.x.

Leite, M.F.A., and Kuramae, E.E. (2020). You must choose, but choose wisely: Model-based approaches for microbial community analysis. *Soil Biol. Biochem.* 151. doi: 10.1016/j.soilbio.2020.108042.

Lemieux, L., and Simard, R., E. (1991). Bitter flavour in dairy products. I. A review of the factors likely to influence its development, mainly in cheese manufacture. *Lait* 71**,** 599-636.

Levante, A., De Filippis, F., La Storia, A., Gatti, M., Neviani, E., Ercolini, D., et al. (2017). Metabolic gene-targeted monitoring of non-starter lactic acid bacteria during cheese ripening. *Int. J. Food Microbiol.* 257**,** 276-284. doi: 10.1016/j.ijfoodmicro.2017.07.002.

Linares, D.M., Del Rio, B., Ladero, V., Martinez, N., Fernandez, M., Martin, M.C., et al. (2012). Factors influencing biogenic amines accumulation in dairy products. *Front. Microbiol.* 3**,** 180. doi: 10.3389/fmicb.2012.00180.

Linares, D.M., Martin, M.C., Ladero, V., Alvarez, M.A., and Fernandez, M. (2011). Biogenic amines in dairy products. *Crit. Rev. Food Sci. Nutr.* 51**,** 691-703. doi: 10.1080/10408398.2011.582813.

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

Mao, F., Leung, W.Y., and Xin, X. (2007). Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol.* 7**,** 76. doi: 10.1186/1472-6750-7-76.

Martin, M.C., Martinez, N., del Rio, B., Ladero, V., Fernandez, M., and Alvarez, M.A. (2010). A novel real-time polymerase chain reaction-based method for the detection and quantification of lactose-fermenting *Enterobacteriaceae* in the dairy and other food industries. *J. Dairy Sci.* 93**,** 860-867. doi: 10.3168/jds.2009-2481.

Martino, C., Morton, J.T., Marotz, C.A., Thompson, L.R., Tripathi, A., Knight, R., et al. (2019). A novel sparse compositional technique reveals microbial perturbations. *mSystems* 4. doi: 10.1128/mSystems.00016-19.

Masoud, W., Vogensen, F.K., Lillevang, S., Abu Al-Soud, W., Sorensen, S.J., and Jakobsen, M. (2012). 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.* 153**,** 192-202. doi: 10.1016/j.ijfoodmicro.2011.11.014.

McSweeney, P.L.H. (2004). "Biochemistry of cheese ripening: Introduction and overview," in *Cheese: Chemistry, Physics and Microbiology,* eds. P.F. Fox, P.L.H. McSweeney, T.M. Cogan & T.P. Guinee. Academic Press**,** 347-360.

McSweeney, P.L.H. (2007). "The microbiology of cheese ripening," in *Cheese Problems Solved,* Woodhead Publishing**,** 117-132.

McSweeney, P.L.H., and Fox, P.F. (2004). "Metabolism of residual lactose and of lactate and citrate," in *Cheese: Chemistry, Physics and Microbiology,* eds. P.F. Fox, P.L.H. McSweeney, T.M. Cogan, T.P. Guinee & Academic Press**,** 361-371.

McSweeney, P.L.H., and Sousa, M.J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. *Lait* 80**,** 293-324.

Mei, J., Guo, Q., Wu, Y., and Li, Y. (2014). Microbial diversity of a Camembert-type cheese using freeze-dried Tibetan kefir coculture as starter culture by culture-dependent and culture-independent methods. *PLoS One* 9**,** e111648. doi: 10.1371/journal.pone.0111648.

Mills, S., Griffin, C., O'Connor, P.M., Serrano, L.M., Meijer, W.C., Hill, C., et al. (2017). A multibacteriocin cheese starter system, comprising nisin and lacticin 3147 in *Lactococcus lactis*, in combination with plantaricin from *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* 83. doi: 10.1128/AEM.00799-17.

Miragoli, F., Patrone, V., Romaniello, F., Rebecchi, A., and Callegari, M.L. (2020). Development of an S-layer gene-based PCR-DGGE assay for monitoring dominant *Lactobacillus helveticus* strains in natural whey starters of Grana Padano cheese. *Food Microbiol.* 89**,** 103457. doi: 10.1016/j.fm.2020.103457.

Moser, A., Berthoud, H., Eugster, E., Meile, L., and Irmler, S. (2017). Detection and enumeration of *Lactobacillus helveticus* in dairy products. *Int. Dairy J.* 68**,** 52-59. doi: 10.1016/j.idairyj.2016.12.007.

Mounier, J., Rea, M.C., O'Connor, P.M., Fitzgerald, G.F., and Cogan, T.M. (2007). Growth characteristics of *Brevibacterium*, *Corynebacterium*, *Microbacterium*, and *Staphylococcus* spp. isolated from surface-ripened cheese. *Appl. Environ. Microbiol.* 73**,** 7732-7739. doi: 10.1128/AEM.01260-07.

Myers, R.M., Maniatis, T., and Lerman, L.S. (1987). "[31] Detection and localization of single base changes by denaturing gradient gel electrophoresis," in *Methods Enzymol.,* Academic Press**,** 501-527.

O'Grady, J., Ruttledge, M., Sedano-Balbas, S., Smith, T.J., Barry, T., and Maher, M. (2009). Rapid detection of *Listeria monocytogenes* in food using culture enrichment combined with real-time PCR. *Food Microbiol.* 26**,** 4-7. doi: 10.1016/j.fm.2008.08.009.

O'Sullivan, D., Ross, R.P., Fitzgerald, G.F., and Coffey, A. (2000). Investigation of the relationship between lysogeny and lysis of *Lactococcus lactis* in cheese using prophage-targeted PCR. *Appl. Environ. Microbiol.* 66**,** 2192-2198.

O'Sullivan, D.J., Cotter, P.D., O'Sullivan, O., Giblin, L., McSweeney, P.L., and Sheehan, J.J. (2015a). Temporal and spatial differences in microbial composition during the manufacture of a continental-type cheese. *Appl. Environ. Microbiol.* 81**,** 2525-2533. doi: 10.1128/AEM.04054-14.

O'Sullivan, D.J., Fallico, V., O'Sullivan, O., McSweeney, P.L., Sheehan, J.J., Cotter, P.D., et al. (2015b). High-throughput DNA sequencing to survey bacterial histidine and tyrosine decarboxylases in raw milk cheeses. *BMC Microbiol.* 15**,** 266. doi: 10.1186/s12866-015-0596-0.

O'Sullivan, D.J., Giblin, L., McSweeney, P.L., Sheehan, J.J., and Cotter, P.D. (2013). Nucleic acid-based approaches to investigate microbial-related cheese quality defects. *Front. Microbiol.* 4**,** 1. doi: 10.3389/fmicb.2013.00001.

Ortakci, F., Broadbent, J.R., Oberg, C.J., and McMahon, D.J. (2015). Growth and gas formation by *Lactobacillus wasatchensis*, a novel obligatory heterofermentative nonstarter lactic acid bacterium, in Cheddar-style cheese made using a *Streptococcus thermophilus* starter. *J. Dairy Sci.* 98**,** 7473-7482. doi: 10.3168/jds.2015-9556.

Oude Elferink, S.J., Krooneman, J., Gottschal, J.C., Spoelstra, S.F., Faber, F., and Driehuis, F. (2001). Anaerobic conversion of lactic acid to acetic acid and 1, 2-propanediol by *Lactobacillus buchneri*. *Appl. Environ. Microbiol.* 67**,** 125-132. doi: 10.1128/AEM.67.1.125-132.2001.

Palles, Beresford, Condon, and Cogan (1998). Citrate metabolism in *Lactobacillus casei* and *Lactobacillus plantarum*. *J. Appl. Microbiol.* 85**,** 147-154. doi: 10.1046/j.1365-2672.1998.00486.x.

Pangallo, D., Sakova, N., Korenova, J., Puskarova, A., Krakova, L., Valik, L., et al. (2014). Microbial diversity and dynamics during the production of May bryndza cheese. *Int. J. Food Microbiol.* 170**,** 38-43. doi: 10.1016/j.ijfoodmicro.2013.10.015.

Parente, E., and Cogan, T.M. (2004). "Starter Cultures: General Aspects," in *Cheese: Chemistry, Physics and Microbiology,* eds. P.F. Fox, P.L.H. McSweeney, T.M. Cogan & T.P. Guinee. Academic Press**,** 123-147.

Parente, E., Guidone, A., Matera, A., De Filippis, F., Mauriello, G., and Ricciardi, A. (2016). Microbial community dynamics in thermophilic undefined milk starter cultures. *Int. J. Food Microbiol.* 217**,** 59-67. doi: 10.1016/j.ijfoodmicro.2015.10.014.

Pega, J., Rizzo, S., Perez, C.D., Rossetti, L., Diaz, G., Ruzal, S.M., et al. (2016). Effect of the addition of phytosterols and tocopherols on *Streptococcus thermophilus* robustness during industrial manufacture and ripening of a functional cheese as evaluated by qPCR and RT-qPCR. *Int. J. Food Microbiol.* 232**,** 117-125. doi: 10.1016/j.ijfoodmicro.2016.06.003.

Perin, L.M., Savo Sardaro, M.L., Nero, L.A., Neviani, E., and Gatti, M. (2017). Bacterial ecology of artisanal Minas cheeses assessed by culture-dependent and -independent methods. *Food Microbiol.* 65**,** 160-169. doi: 10.1016/j.fm.2017.02.005.

Pircher, A., Bauer, F., and Paulsen, P. (2007). Formation of cadaverine, histamine, putrescine and tyramine by bacteria isolated from meat, fermented sausages and cheeses. *Eur. Food Res. Technol.* 226**,** 225-231. doi: 10.1007/s00217-006-0530-7.

Pogacic, T., Mancini, A., Santarelli, M., Bottari, B., Lazzi, C., Neviani, E., et al. (2013). 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.* 36**,** 207-215. doi: 10.1016/j.fm.2013.05.009.

Porcellato, D., and Skeie, S.B. (2016). Bacterial dynamics and functional analysis of microbial metagenomes during ripening of Dutch-type cheese. *Int. Dairy J.* 61**,** 182-188. doi: 10.1016/j.idairyj.2016.05.005.

Props, R., Kerckhof, F.M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., et al. (2017). Absolute quantification of microbial taxon abundances. *ISME J* 11**,** 584-587. doi: 10.1038/ismej.2016.117.

Quiberoni, A., Guglielmotti, D., and Reinheimer, J. (2008). New and classical spoilage bacteria causing widespread blowing in Argentinean soft and semihard cheeses. *Int. J. Dairy Technol.* 61**,** 358-363. doi: 10.1111/j.1471-0307.2008.00431.x.

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

Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., and Cotter, P.D. (2011). Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese. *Int. J. Food Microbiol.* 150**,** 81-94. doi: 10.1016/j.ijfoodmicro.2011.08.001.

Quigley, L., O'Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., and Cotter, P.D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Appl. Environ. Microbiol.* 78**,** 5717-5723. doi: 10.1128/AEM.00918-12.

Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T.P., Ross, R.P., Fitzgerald, G.F., et al. (2013). The complex microbiota of raw milk. *FEMS Microbiol. Rev.* 37**,** 664-698. doi: 10.1111/1574-6976.12030.

Quinn, T.P., Erb, I., Gloor, G., Notredame, C., Richardson, M.F., and Crowley, T.M. (2019). A field guide for the compositional analysis of any-omics data. *Gigascience* 8. doi: 10.1093/gigascience/giz107.

Ramezani, M., Hosseini, S.M., Ferrocino, I., Amoozegar, M.A., and Cocolin, L. (2017). Molecular investigation of bacterial communities during the manufacturing and ripening of semi-hard Iranian Liqvan cheese. *Food Microbiol.* 66**,** 64-71. doi: 10.1016/j.fm.2017.03.019.

Rapposch, S., Eliskases-Lechner, F., and Ginzinger, W. (1999). Growth of facultatively heterofermentative lactobacilli on starter cell suspensions. *Appl. Environ. Microbiol.* 65**,** 5597-5599.

Rossetti, L., Fornasari, M.E., Gatti, M., Lazzi, C., Neviani, E., and Giraffa, G. (2008). Grana Padano cheese whey starters: microbial composition and strain distribution. *Int. J. Food Microbiol.* 127**,** 168-171. doi: 10.1016/j.ijfoodmicro.2008.06.005.

Rossi, F., Gardini, F., Rizzotti, L., La Gioia, F., Tabanelli, G., and Torriani, S. (2011). Quantitative analysis of histidine decarboxylase gene (*hdcA*) transcription and histamine production by *Streptococcus thermophilus* PRI60 under conditions relevant to cheese making. *Appl. Environ. Microbiol.* 77**,** 2817-2822. doi: 10.1128/AEM.02531-10.

Rossmanith, P., Krassnig, M., Wagner, M., and Hein, I. (2006). Detection of *Listeria monocytogenes* in food using a combined enrichment/real-time PCR method targeting the *prfA* gene. *Res. Microbiol.* 157**,** 763-771. doi: 10.1016/j.resmic.2006.03.003.

Roth, E., Miescher Schwenninger, S., Hasler, M., Eugster-Meier, E., and Lacroix, C. (2010). Population dynamics of two antilisterial cheese surface consortia revealed by temporal temperature gradient gel electrophoresis. *BMC Microbiol.* 10**,** 74. doi: 10.1186/1471-2180-10-74.

Ruggirello, M., Cocolin, L., and Dolci, P. (2016). Fate of *Lactococcus lactis* starter cultures during late ripening in cheese models. *Food Microbiol.* 59**,** 112-118. doi: 10.1016/j.fm.2016.05.001.

Ruggirello, M., Dolci, P., and Cocolin, L. (2014). Detection and viability of *Lactococcus lactis* throughout cheese ripening. *PLoS One* 9**,** e114280. doi: 10.1371/journal.pone.0114280.

Ruggirello, M., Giordano, M., Bertolino, M., Ferrocino, I., Cocolin, L., and Dolci, P. (2018). Study of *Lactococcus lactis* during advanced ripening stages of model cheeses characterized by GC-MS. *Food Microbiol.* 74**,** 132-142. doi: 10.1016/j.fm.2018.03.012.

Ryssel, M., Johansen, P., Al-Soud, W.A., Sorensen, S., Arneborg, N., and Jespersen, L. (2015). 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.* 215**,** 124-130. doi: 10.1016/j.ijfoodmicro.2015.09.012.

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230**,** 1350-1354. doi: 10.1126/science.2999980.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., et al. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239**,** 487-491. doi: 10.1126/science.2448875.

Salvetti, E., Harris, H.M.B., Felis, G.E., and O'Toole, P.W. (2018). Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl. Environ. Microbiol.* 84. doi: 10.1128/AEM.00993-18.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74**,** 5463-5467. doi: 10.1073/pnas.74.12.5463.

SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences* 95**,** 1460-1465. doi: 10.1073/pnas.95.4.1460.

Santarelli, M., Bottari, B., Lazzi, C., Neviani, E., and Gatti, M. (2013). Survey on the community and dynamics of lactic acid bacteria in Grana Padano cheese. *Syst. Appl. Microbiol.* 36**,** 593-600. doi: 10.1016/j.syapm.2013.04.007.

Santarelli, M., Gatti, M., Lazzi, C., Bernini, V., Zapparoli, G.A., and Neviani, E. (2008). Whey starter for Grana Padano cheese: effect of technological parameters on viability and composition of the microbial community. *J. Dairy Sci.* 91**,** 883-891. doi: 10.3168/jds.2007-0296.

Sartori, C., Boss, R., Ivanovic, I., and Graber, H.U. (2017). Development of a new real-time quantitative PCR assay for the detection of *Staphylococcus aureus* genotype B in cow milk, targeting the new gene *adlb*. *J. Dairy Sci.* 100**,** 7834-7845. doi: 10.3168/jds.2017-12820.

Savo Sardaro, M.L., Perin, L.M., Bancalari, E., Neviani, E., and Gatti, M. (2018). Advancement in LH-PCR methodology for multiple microbial species detections in fermented foods. *Food Microbiol.* 74**,** 113-119. doi: 10.1016/j.fm.2018.03.008.

Schmitz-Esser, S., Dzieciol, M., Nischler, E., Schornsteiner, E., Bereuter, O., Mann, E., et al. (2018). Abundance and potential contribution of Gram-negative cheese rind bacteria from Austrian artisanal hard cheeses. *Int. J. Food Microbiol.* 266**,** 95-103. doi: 10.1016/j.ijfoodmicro.2017.11.013.

Settanni, L., and Moschetti, G. (2010). Non-starter lactic acid bacteria used to improve cheese quality and provide health benefits. *Food Microbiol.* 27**,** 691-697. doi: 10.1016/j.fm.2010.05.023.

Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmler, S., et al. (2019). Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiol.* 19**,** 143. doi: 10.1186/s12866-019-1500-0.

Sun, Z., Harris, H.M., McCann, A., Guo, C., Argimon, S., Zhang, W., et al. (2015). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat Commun* 6**,** 8322. doi: 10.1038/ncomms9322.

Thierry, A., Collins, Y.F., Abeijón Mukdsi, M.C., McSweeney, P.L.H., Wilkinson, M.G., and Spinnler, H.E. (2017). "Lipolysis and metabolism of fatty acids in cheese," in *Cheese***,** 423-444.

Thomas, T.D. (1987). Acetate production from lactate and citrate by nonstarter bacteria in Cheddar cheese. *New Zealand Journal of Dairy Science and Technology* 22**,** 25-38.

Treimo, J., Vegarud, G., Langsrud, T., and Rudi, K. (2006). Use of DNA quantification to measure growth and autolysis of *Lactococcus* and *Propionibacterium* spp. in mixed populations. *Appl. Environ. Microbiol.* 72**,** 6174-6182. doi: 10.1128/AEM.00515-06.

Tsakalidou, E., Manolopoulou, E., Kabaraki, E., Zoidou, E., Pot, B., Kersters, K., et al. (1994). The combined use of whole-cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from traditional Greek dairy products. *Syst. Appl. Microbiol.* 17**,** 444-458. doi: 10.1016/s0723-2020(11)80062-7.

Turgay, M., Schaeren, W., Graber, H.U., Wagner, E., Amrein, R., Bütikofer, U., et al. (2018). A field study investigating the effectiveness of vat milk controls by qPCR for the prevention of undesired propionic acid fermentation in Sbrinz PDO cheese. *Int. Dairy J.* 77**,** 80-88. doi: 10.1016/j.idairyj.2017.06.008.

Turgay, M., Schaeren, W., Wechsler, D., Bütikofer, U., and Graber, H.U. (2016). Fast detection and quantification of four dairy propionic acid bacteria in milk samples using real-time quantitative polymerase chain reaction. *Int. Dairy J.* 61**,** 37-43. doi: 10.1016/j.idairyj.2016.03.014.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., et al. (2012). Primer3–new capabilities and interfaces. *Nucleic Acids Res.* 40**,** e115. doi: 10.1093/nar/gks596.

Upadhyay, V.K., McSweeney, P.L.H., Magboul, A.A.A., and Fox, P.F. (2004). "Proteolysis in Cheese during Ripening," in *Cheese: Chemistry, Physics and Microbiology,* eds. P.F. Fox, P.L.H. McSweeney, T.M. Cogan & T.P. Guinee. Academic Press, 1(3)**,** 391-434.

Valence, F., Richoux, R., Thierry, A., Palva, A., and Lortal, S. (1998). Autolysis of *Lactobacillus helveticus* and *Propionibacterium freudenreichii* in Swiss cheeses: first evidence by using species-specific lysis markers. *Journal of Dairy Research (United Kingdom).*

Visweswaran, G.R., Kurek, D., Szeliga, M., Pastrana, F.R., Kuipers, O.P., Kok, J., et al. (2017). Expression of prophage-encoded endolysins contributes to autolysis of *Lactococcus lactis*. *Appl. Microbiol. Biotechnol.* 101**,** 1099-1110. doi: 10.1007/s00253-016-7822-z.

von Ahsen, N., Wittwer, C.T., and Schütz, E. (2001). Oligonucleotide melting temperatures under PCR conditions: nearest-neighbor corrections for Mg(2+), deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin. Chem.* 47**,** 1956-1961.

Wilhelm, J., and Pingoud, A. (2003). Real-time polymerase chain reaction. *ChemBioChem* 4**,** 1120-1128. doi: 10.1002/cbic.200300662.

Wolfe, B.E., Button, J.E., Santarelli, M., and Dutton, R.J. (2014). Cheese rind communities provide tractable systems for *in situ* and *in vitro* studies of microbial diversity. *Cell* 158**,** 422-433. doi: 10.1016/j.cell.2014.05.041.

Xue, Z., Brooks, J.T., Quart, Z., Stevens, E.T., Kable, M.E., Heidenreich, J., et al. (2021). Microbiota assessments for the identification and confirmation of slit defect-causing bacteria in milk and Cheddar cheese. *mSystems* 6**,** e01114-01120. doi: 10.1128/mSystems.01114-20.

Yang, B., Wang, Y., and Qian, P.Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 17**,** 135. doi: 10.1186/s12859-016-0992-y.

Yang, C., You, L., Kwok, L.-Y., Jin, H., Peng, J., Zhao, Z., et al. (2021). Strain-level multiomics analysis reveals significant variation in cheeses from different regions. *Lwt* 151. doi: 10.1016/j.lwt.2021.112043.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13**,** 134. doi: 10.1186/1471-2105-13-134.

Yunita, D., and Dodd, C.E.R. (2018). Microbial community dynamics of a blue-veined raw milk cheese from the United Kingdom. *J. Dairy Sci.* 101**,** 4923-4935. doi: 10.3168/jds.2017-14104.

Yvon, M., and Rijnen, L. (2001). Cheese flavour formation by amino acid catabolism. *Int. Dairy J.* 11**,** 185-201. doi: 10.1016/S0958-6946(01)00049-8.

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

Zheng, J., Wittouck, S., Salvetti, E., Franz, C., Harris, H.M.B., Mattarelli, P., et al. (2020). A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int. J. Syst. Evol. Microbiol.* doi: 10.1099/ijsem.0.004107.

Zuker, M., Mathews, D.H., and Turner, D.H. (1999). "Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide," in *RNA Biochemistry and Biotechnology,* eds. J. Barciszewski & B.F.C. Clark. (Dordrecht: Springer Netherlands)**,** 11-43.

Chapter 2

SpeciesPrimer: a bioinformatics pipeline dedicated to the design of qPCR primers for the quantification of bacterial species

PeerJ

SpeciesPrimer: a bioinformatics pipeline dedicated to the design of qPCR primers for the quantification of bacterial species

Matthias Dreier[1,2], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1] and Pilar Junier[2]

[1] Agroscope, Bern, Switzerland
[2] Laboratory of Microbiology, University of Neuchâtel, Neuchâtel, Switzerland

ABSTRACT

Background. Quantitative real-time PCR (qPCR) is a well-established method for detecting and quantifying bacteria, and it is progressively replacing culture-based diagnostic methods in food microbiology. High-throughput qPCR using microfluidics brings further advantages by providing faster results, decreasing the costs per sample and reducing errors due to automatic distribution of samples and reagents. In order to develop a high-throughput qPCR approach for the rapid and cost-efficient quantification of microbial species in complex systems such as fermented foods (for instance, cheese), the preliminary setup of qPCR assays working efficiently under identical PCR conditions is required. Identification of target-specific nucleotide sequences and design of specific primers are the most challenging steps in this process. To date, most available tools for primer design require either laborious manual manipulation or high-performance computing systems.

Results. We developed the SpeciesPrimer pipeline for automated high-throughput screening of species-specific target regions and the design of dedicated primers. Using SpeciesPrimer, specific primers were designed for four bacterial species of importance in cheese quality control, namely *Enterococcus faecium*, *Enterococcus faecalis*, *Pediococcus acidilactici* and *Pediococcus pentosaceus*. Selected primers were first evaluated *in silico* and subsequently *in vitro* using DNA from pure cultures of a variety of strains found in dairy products. Specific qPCR assays were developed and validated, satisfying the criteria of inclusivity, exclusivity and amplification efficiencies.

Conclusion. In this work, we present the SpeciesPrimer pipeline, a tool to design species-specific primers for the detection and quantification of bacterial species. We use SpeciesPrimer to design qPCR assays for four bacterial species and describe a workflow to evaluate the designed primers. SpeciesPrimer facilitates efficient primer design for species-specific quantification, paving the way for a fast and accurate quantitative investigation of microbial communities.

Submitted 10 July 2019
Accepted 10 January 2020
Published 18 February 2020

Corresponding author
Matthias Dreier,
matthias.dreier@agroscope.admin.ch,
matthias.dreier@unine.ch

Academic editor
Cindy Smith

Additional Information and Declarations can be found on page 21

DOI **10.7717/peerj.8544**

Subjects Bioinformatics, Food Science and Technology, Microbiology

Keywords Primer design, Species specific quantification, Quantitative real-time polymerase chain reaction, qPCR primer, Species specific sequences, Docker container, Bioinformatics pipeline, Primer validation

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

PeerJ

INTRODUCTION

Quantitative real-time PCR (qPCR) is a well-established method for the detection and quantification of bacteria in microbiology, for instance in the context of pathogen detection in clinical and veterinary diagnostics and food safety (*Cremonesi et al., 2014*; *Curran et al., 2007*; *Garrido-Maestu et al., 2018*; *Ramirez et al., 2009*). Culture-based diagnostic methods are progressively being replaced by qPCR due to advantages such as faster results, more specific detection, and the ability to detect sub-dominant populations (*Postollec et al., 2011*). High-throughput microfluidic qPCR brings further advantages including the fast generation of results, a lower cost per sample and fewer errors due to automatic distribution of samples and reagents. However, in order to work efficiently, high-throughput qPCR systems use identical PCR chemistry and PCR conditions for all reactions taking place on a single chip. Therefore, existing qPCR assays are often not suitable and new primers have to be designed (*Hermann-Bank et al., 2013*; *Ishii, Segawa & Okabe, 2013*; *Kleyer, Tecon & Or, 2017*).

The main challenges for the successful development of any qPCR assay are the identification of a specific target nucleotide sequence and the design of primers that bind exclusively to that target sequence. Before microbial draft genomes became widely available, the 16S rRNA gene was frequently used as a target sequence. However, the regions that are targeted in the 16S rRNA gene often do not provide sufficient resolution to differentiate between closely related bacterial species (*Moyaert et al., 2008*; *Torriani, Felis & Dellaglio, 2001*; *Wang et al., 2007*). Further, housekeeping genes such as, for instance, *tuf*, *recA* and *pheS*, were successfully used as target sequences for a variety of bacterial species in fermented foods (*Falentin et al., 2010*; *Masco et al., 2007*; *Scheirlinck et al., 2009*). Today, the steadily increasing number of prokaryotic draft genomes facilitates the identification of new and unique target regions. This, in combination with the increased computing power, makes it now possible to screen and compare hundreds of genomes and to predict unique target sequences in a relatively short time.

Various commercial and open source programs facilitate the design of specific primers for a target sequence, such as the standard tools Primer3 and Primer-BLAST (*Untergasser et al., 2012*; *Ye et al., 2012*). Primer3 predicts suitable PCR primers for an input target sequence, while Primer-BLAST combines Primer3 with a BLAST search in a selected nucleotide sequence database to assess the specificity of the primers for the target sequence. Table 1 provides an overview of the features of different primer design tools and pipelines. PrimerMiner (*Elbrecht, Leese & Bunce, 2017*) is a tool that automatically downloads sequences of marker genes for taxonomic groups specified by the user and creates alignments and consensus sequences as target sequences for the design of degenerate primers. PrimerServer (*Zhu et al., 2017*) allows to design primers for multiple sites across a whole genome sequence and performs a specificity check. Tools and pipelines that encompass both the identification of target sequences from bacterial draft genomes and the design of primer candidates include, for instance, RUCS, the find_differential_primers (fdp) pipeline and TOPSI (*Pritchard et al., 2012*; *Thomsen et al., 2017*; *Vijaya Satya et al., 2010*). RUCS is able to identify unique core sequences in a positive set of genomes

(target) compared to a negative set of genomes (non-target). It designs primers for the core sequences and validates them with an *in silico* PCR validation method against the positive and negative reference sets. Similarly, the fdp pipeline designs primers for a set of positive genomes and, further, allows to extract primers specific to subclasses of the positive set and performs specificity check against a negative set of genomes. TOPSI is an automated high-throughput pipeline for the design of primers, primarily developed for pathogen-diagnostic assays. It identifies sequences present in all input genomes and designs specific primers accordingly.

We aimed to design a series of primers that function with the same qPCR cycling conditions and primer concentrations for later usage in a high-throughput microfluidic qPCR platform. RUCS, fdp and TOPSI can be used to design species-specific primers and offer high-throughput primer design. However, TOPSI could not be used because no Linux-based cluster was available. RUCS and fdp were initially not able to design primers for all our target species. Therefore, these pipelines were not suitable for our high-throughput approach.

This study presents a new pipeline named SpeciesPrimer developed for automated high-throughput screening for species-specific target regions combined with the design of primer candidates for these sequences. The process of primer design is fully automated from the download of bacterial genomes to the quality control of primer candidates. The pipeline runs on a standard computer with a multi-core processor and a minimum of 16 GB RAM. We have applied the SpeciesPrimer pipeline to a set of four bacterial species occurring in cheese and other dairy products and validated the primers *in silico* and *in vitro* by performing qPCR experiments with a variety of target and non-target strains.

DESCRIPTION

Overview

The SpeciesPrimer pipeline consists of three main parts (Table 2). First, genome assemblies are downloaded, annotated and then subjected to quality control. Second, a pan-genome analysis is performed to identify single copy core genes. Conserved sequences of these core genes are then extracted and the specificity for the target species is assessed. Finally, primers are designed for these species-specific conserved core gene sequences and subsequently evaluated in a primer quality control step. An overview of the features of the tools used for SpeciesPrimer can be found in Table S1.

Part 1: Input genome assemblies

The minimal command line input for the pipeline is the species name. Further, a list of non-target species names can be specified (e.g., species found in the investigated ecosystem but that should not be detected in the specific qPCR assay). For downloading genome assemblies from the National Center for Biotechnology Information (NCBI) automatically, a valid e-mail address is required for accessing the NCBI E-utilities services (*Sayers, 2009*). The pipeline works with a pre-formatted NCBI BLAST database (nt), containing partially non-redundant nucleotide sequences. A local copy of the nt database is required. It can be downloaded from NCBI using the update_blastdb.pl script from

Table 1 Overview of the features of different primer design tools and pipelines.

Tool Reference	RUCS *Thomsen et al. (2017)*	fdp *Pritchard et al. (2012)*	TOPSI *Vijaya Satya et al. (2010)*	Species-Primer (this study)	Primer-Miner *Elbrecht, Leese & Bunce (2017)*	Primer-Server *Zhu et al. (2017)*	Primer-BLAST *Ye et al. (2012)*
Primer specificity	Bacterial strains / species		Bacterial species		Taxonomic groups	Input sequence	
Inputs							
Taxonomic group(s)	–	–	–	Species	Order, Family	–	–
Target gene(s)	–	–	–	–	x	–	x
Genome assemblies	x	x	x	x	–	–	–
Target sequences	–	–	–	–	x	x	x
Primer sequences	x	x	–	–	x	x	x
Automatic download of input sequences	–	–	–	x	x	–	–
Identification of target sequences	x	x	x	x	–	–	–
Identification of conserved regions	x	–	x	x	x	–	–
Primer design	x	x	x	x	–	x	x
Specificity check							
Target sequences	Input sequences	Input sequences	BLAST DB	BLAST DB	–	–	BLAST DB
Primer	Input sequences	BLAST DB	BLAST DB	BLAST DB	Alignment	BLAST DB	BLAST DB
Primer quality control	x	x	x	x	–	x	
Primer3 cutoffs	x	x	x	x	–	x	x
Primer dimer	–	–	–	x	–	–	–
Hairpin	–	–	–	–	–	–	–
Amplicon secondary structures	–	–	–	x	–	–	–

(*continued on next page*)

Table 1 (*continued*)

Tool Reference	RUCS *Thomsen et al. (2017)*	fdp *Pritchard et al. (2012)*	TOPSI *Vijaya Satya et al. (2010)*	Species-Primer (this study)	Primer-Miner *Elbrecht, Leese & Bunce (2017)*	Primer-Server *Zhu et al. (2017)*	Primer-BLAST *Ye et al. (2012)*
High-throughput primer design	x	x	x	x	–	x	–
Batch processing	–	–	–	Full runs	Download	–	–
Works on standard computers	x	x	–	x	x	x	–
Graphic user interface	x	–	–	x	–	x	x
Web service	x	–	*	–	–	x	x

Notes.
fdp, find_differential_primers; x, Feature supported; –, Feature not supported; *, Access has to be requested; QC, Quality control; CDS, Coding sequences.

Table 2 Overview of the SpeciesPrimer pipeline workflow and the used software.

Pipeline workflow	Tools (Version[a])	Reference
Input genome assemblies		
- download	NCBI Entrez (Biopython 1.73)	*Cock et al. (2009)*, *Sayers (2009)*
- annotation	Prokka (1.13.7)	*Seemann (2014)*
- quality control	BLAST+ (2.9.0+)	*Altschul et al. (1990)*
Core gene sequences		
- identification	Roary (3.12.0)	*Page et al. (2015)*
- phylogeny	FastTree 2 (2.1.11)	*Price, Dehal & Arkin (2010)*
- selection of conserved sequences	Prank (.150803) consambig (EMBOSS 6.6.0.0) GNU parallel (20161222)	*Löytynoja (2014)* *Rice, Longden & Bleasby (2000)* *Tange (2011)*
- evaluation of specificity	BLAST+	*Altschul et al. (1990)*
Primer		
- design	Primer3 (2.4.0)	*Untergasser et al. (2012)*
- quality control	BLAST+, MFEprimer (2.0), MPprimer (1.5), Mfold (3.6)	*Altschul et al. (1990)* *Qu et al. (2012)* *Shen et al. (2010)* *Zuker, Mathews & Turner (1999)*

Notes.
[a]Docker image June 13, 2019.

the BLAST+ package (*Altschul et al., 1990*), via FTP from the NCBI FTP server or with the pipeline script (getblastdb.py). The nt database, which consists of sequences from GenBank, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Data Bank of Japan), was selected because it has a large coverage of diverse sequences, but it is not as large as for example the refseq_genomic database (*Tao et al., 2011*). The evaluation of the specificity of the target sequence for the target species does not rely on small differences in the nucleotide sequence, but on the overall similarity. Therefore, even with one genome sequence per non-target species we would expect to find similarities in the core genes of the non-target species. Each additional genome of this species in a database would then allow finding more potential sequence similarities in shell genes, cloud genes and strain-specific genes. On the one hand, a more extensive database could better predict the specificity of a sequence, but on the other, it would increase the size of the database and the time required for the BLAST search.

The user-provided species name is used to search for genome assemblies in the NCBI database. The Biopython Entrez module (*Cock et al., 2009*) searches the NCBI taxonomic identity (taxid) for the target species in the taxonomy database and downloads the genome assembly summary report. Afterwards, SpeciesPrimer downloads the genome assemblies in FASTA format from the NCBI RefSeq FTP server using the links specified in the summary report. Finally, the downloaded genome assemblies are annotated with Prokka (*Seemann, 2014*).

The quality of the genome assemblies is a crucial factor for the pan-genome analysis. Genome assemblies deposited with the wrong taxonomic label or low-quality assemblies drastically reduce the number of identified core genes and of conserved sequences for

primer design. The initial quality control step is intended to remove such assemblies from the subsequent analysis. For the verification of the taxonomic classification, the user can choose one or several genes from five conserved housekeeping genes (16S rRNA, *tuf, recA, dnaK* and *pheS*). Genome assemblies without an annotation for the specified conserved housekeeping genes or genome assemblies consisting of more than 500 contigs are removed from the downstream pan-genome analysis. The sequences of the specified conserved housekeeping genes are blasted against the local nt database. Genome assemblies pass the quality control if the best BLAST hit for all sequences is a sequence arising from the target species.

Part 2: Identification of target sequences for primer design

A pan-genome analysis is performed using Roary (*Page et al., 2015*) to identify the core genes of the target species. Based on the results of the pan-genome analysis, single copy core genes are identified. The gene_presence_absence.csv produced by Roary reports the presence (or absence) of every annotated gene for every input genome assembly. Single copy core genes are the genes for which the number of assemblies harboring the sequence and the number of total identified sequences equals the number of total input assemblies. An sqlite3 database containing all annotated sequences of all assemblies is compiled using the DBGenerator.py script from the Microbial Genomics Lab GitHub repository (https://github.com/microgenomics/tutorials). This database is queried for single copy core genes and the nucleotide sequences are saved in multi-FASTA format. Each multi-FASTA file contains the sequences of one single copy core gene from each input genome assembly. These sequences are aligned using the probabilistic multiple alignment program Prank (*Löytynoja, 2014*). A consensus sequence with ambiguous bases is then created using the consambig function from the EMBOSS package (*Rice, Longden & Bleasby, 2000*). The alignments and extraction of the consensus sequences are performed in parallel for several core genes using GNU parallel (*Tange, 2011*). Continuous consensus sequences longer than the minimal PCR product length, harboring less than two ambiguous bases in the range of 20 bases are used for the subsequent steps of the pipeline.

These conserved consensus sequences are used for a BLAST search against the local nt database using the discontiguous BLAST algorithm and an *e*-value cutoff of 500. For all hits in the BLAST results, the species name is extracted from the sequence description and compared with the names in the species list (non-target species). If any species name in the species list matches a hit in the BLAST results the corresponding query sequence is discarded, otherwise the sequence is classified as specific for the target and considered for primer design.

Part 3: Primer design

Primer3 is used to design primers for the unique single copy core gene sequences. As pipeline default, the optimum primer melting temperature is set to 60 °C and the maximal primer length is set to 26 bases. All other settings are the default settings of the primer3web version (http://primer3.ut.ee, accessed November 29, 2018). The minimal and maximal amplicon size of the PCR product can be specified individually for every target species

through the command line options. The other parameters for Primer3 cannot be changed individually, but the general Primer3 settings can be changed by modifying the Primer3 settings file.

The primer quality control consists of three parts, an *in silico* PCR to evaluate the specificity of the primer for the template, an estimation of secondary structures of the amplicon sequence and the estimation of the potential to form primer dimers. The specificity check (Fig. 1) for each primer pair is performed with MFEprimer-2.0 (*Qu et al., 2012*). For the evaluation of the specificity, three indexed databases are generated: the target template database, the non-target sequence database and the target genome database. The target template database consists of the unique conserved core gene sequences used as template for primer design. The non-target sequence database is compiled from sequences of non-target species that show similarities to the primer sequences. To identify these sequences, a BLAST search with all primers against the local nt database is performed. BLAST hits with a species name in the description matching a name in the user-specified non-target species list are selected. These selected sequences and 4000 base pairs up- and downstream are extracted from the nt database using the blastdbcmd tool. The target genome database is composed of maximal 10 of the input genome assemblies. If the assembly summary report from the automatic download of genome assemblies from NCBI is available, the genome assemblies as complete as possible are preferred (assembly status: complete >chromosome >scaffold >contig). The target sequence database is used to evaluate the maximum primer pair coverage (PPC, maximum value = 100), a value used by MFEprimer-2.0 to score the ability of the primer pair to bind to a DNA template. All primer pairs with a PPC value lower than the specified threshold (mfethreshold, default = 90) for their template are excluded. Next, MFEprimer-2.0 is used to score the binding of the primer pairs to the sequences of the non-target sequence and the target genome database. The difference of the PPC for the DNA template and the specified threshold (Δthreshold = PPC –mfethreshold) is used as a threshold for the maximum PPC value a primer pair is allowed to have for a non-target sequence. Strong secondary structures at the 5′- or the 3′- end of the PCR product could impair efficient primer binding. Therefore, the PCR products of the primer pairs are submitted to mfold (*Zuker, Mathews & Turner, 1999*) to exclude PCR products with strong secondary structures at the annealing temperature of 60 °C. Moreover, as primer dimers can yield unspecific signals during the qPCR run, the 3′- ends of the primer pairs are checked for their potential to form homo- or hetero-dimers using a Perl script (MPprimer_dimer_check.pl) from MPprimer (*Shen et al., 2010*).

The pipeline output is a list containing the primer name, primer pair coverage (MFEprimer) and penalty values, primer and template sequences and melting temperatures (Primer3). Further, a report of the genome assembly quality control, a file containing the pipeline run statistics, the core gene alignment and the phylogeny in newick format can be found in the output directory.

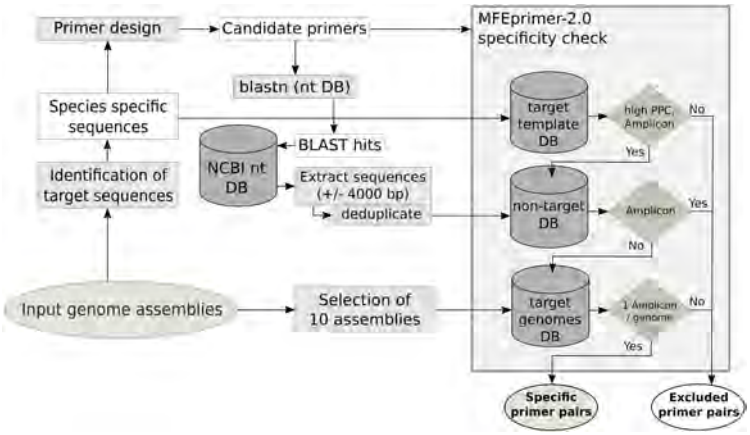


Figure 1 **Schematic workflow of the database creation and the specificity check using MFEprimer-2.0.**
Full-size DOI: 10.7717/peerj.8544/fig-1

MATERIALS & METHODS

Primer design

SpeciesPrimer pipeline runs were performed on a virtual machine (Oracle VM VirtualBox 5.2.8) with Ubuntu 16.04 (64-bit) and Docker installed, using 22 of 24 logical processors from two Intel Xeon E5-2643 CPUs, 32 GB of RAM, a solid-state drive and a LAN Internet connection. To show the performance of the SpeciesPrimer pipeline on consumer hardware, the runs were repeated on a laptop with an Ubuntu 16.04 (64-bit) operating system, an i7-3610QM CPU (8 logical processors), 8 GB RAM, a solid-state drive and a wireless LAN Internet connection. The used Docker image is available at https://hub.docker.com/r/biologger/speciesprimer.

The species list consisted of 259 species and subspecies names detected in dairy products, namely from species names collected from data of 16S rRNA amplicon sequencing studies in milk and cheese varieties (Marco Meola Agroscope, pers. comm.) and dairy-related bacteria from the list of bacterial species and subspecies with technological beneficial use in food products (*Almeida et al., 2014*).

The SpeciesPrimer pipeline was run with the input genome assemblies, parameters and the species list specified in the supplemental information (Dataset S1). Genome assemblies from the Agroscope Culture Collection were included for the Pediococci.

In silico validation

For the *in silico* validation, PCR products for the designed primer pairs were used for an online BLAST search against the RefSeq Genomes Database (refseq_genomes) limited to bacterial genomes. The search was performed by qblast (Biopython), using blastn, the maximum hitlist size was set to 5000 and the expect threshold (*e*-value) was set to 500.

Primer pairs were tested for specificity using online Primer-BLAST. The primers were blasted against the nucleotide collection BLAST database (nr) limited to sequences from

bacteria. The nr (non-redundant nucleotide) database was chosen to get the broadest coverage for the BLAST search. Default settings were used, except for the primer specificity stringency that was set to ignore targets that have nine or more mismatches to the primer.

In vitro validation

The inclusivity of the primer pairs was assayed by performing qPCR with 2 ng DNA of 21 to 25 strains of the target species in technical duplicates. The PCR efficiency was examined by ten-fold dilution series of the type strain DNA in a range from 10^6 to 10^1 genome copies per reaction. DNA concentration for the corresponding number of genome copies was estimated by taking the genome size of the type strain (https://www.ncbi.nlm.nih.gov/genome) and an average weight of $1.096 \cdot 10^{-21}$ g per base pair.

The exclusivity of the primer pairs was assayed by performing qPCR in technical duplicates with 2 ng DNA from various bacterial species found in dairy products. Because the number of samples per run was limited, four separate runs were required to measure all non-target strains. In each run three strains of the target species (positive control) and a no template control were included.

Bacterial strains

Strains stored within the Agroscope Culture Collection at −80 °C in sterile reconstituted skim milk powder (10%, w/v), were reactivated and cultivated according to the conditions specified in Dataset S2.

DNA extraction

Unless otherwise noted, all reagents were purchased from Merck, Darmstadt, Germany.

Bacterial pellets harvested from 1 ml culture by centrifugation (10,000× g, 5 min, room temperature) were used for DNA extraction. For a pre-lysis treatment, the bacterial cells were incubated in 1 ml of 50 mM sodium hydroxide for 15 min at room temperature. Afterwards cells were collected by centrifugation (10,000× g, 5 min, room temperature) and then treated with lysozyme (2.5 mg/ml dissolved in 100 mM Tris(hydroxymethyl)aminomethane, 10 mM ethylendiaminetetraacetic acid (EDTA; Calbiochem, San Diego, USA), 25% (w/v) sucrose, pH 8.0) for 1 h at 37 °C. After the pre-lysis treatment, the bacterial cells were collected by centrifugation (10,000× g, 5 min, room temperature). Cell lysis and genomic DNA extraction was performed using the EZ1 DNA Tissue kit and a BioRobot® EZ1 workstation (Qiagen, Hilden, Germany) according to the manufacturer's instructions and eluted in a volume of 100 µl. The DNA concentration was measured using a NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific, Waltham, MA, USA).

Quantitative real-time PCR

The qPCR assays were performed in a total reaction mix volume of 12 µl containing 6 µl 2x SsoFast™ EvaGreen® Supermix with low ROX (Biorad, Cressier, Switzerland), 500 nM of forward and reverse primers, respectively, and 2 µl of DNA. Each sample was measured in technical duplicates. The qPCR cycling conditions were an initial denaturation at 95 °C

for 1 min followed by 35 cycles of 95 °C for 5 s and 60 °C for 1 min. For the melting curve analysis, a gradient from 60–95 °C with 1 °C steps per 3 s was performed. All qPCR assays were run on a Corbett Rotor-Gene 3000 (Qiagen). The analysis was performed using Rotor-Gene 6000 Software 1.7 with dynamic tube normalization and a threshold of 0.05 for quantification cycle (Cq) value calculation, the five first cycles were ignored for the determination of the Cq values. The peak calling threshold for the melt curve analysis was set to -2 dF/dT and a temperature threshold was set 2 °C lower than the positive control peak.

Phylogeny and average nucleotide identity calculations

The phylogeny was created with Roary and FastTree 2 during the pipeline runs and iTOL (*Letunic & Bork, 2019*) was used to visualize the tree. Average nucleotide identity (ANI) calculations were performed with pyani v0.2.9 (*Pritchard et al., 2016*) using the ANIm method. The heatmap was created from the ANIm_percentage_identity.tab output file using the clustermap function of the python seaborn module and modified color bar settings from pyani. For the color bars on top and on the left of the heatmap, the assemblies were assigned to the same color as in the phylogeny tree. Row and column names (genome assembly accessions) can be found in Dataset S3.

Comparison of primer design pipelines

The positive genome sets for RUCS and fdp were the same genome assemblies used for the SpeciesPrimer pipeline. SpeciesPrimer uses by default the NCBI nt database and the species list for the specificity checks, whereas RUCS and fdp require a negative set of genomes. Therefore, a set of (representative) genome assemblies from NCBI was downloaded for the species from the species list. From these assemblies a BLAST database was prepared for SpeciesPrimer. The same genome assemblies, excluding the assembly of the target species, were used as a negative set for RUCS and to make a BLAST database for fdp. For both tools, the minimal and maximal PCR product size was set to 70 and 200, respectively. The tab separated config file for fdp was created using the assembly accession as name, the species as class and providing the absolute path of the genome assembly files. The script was started with the blastdb option to provide the path to the previously prepared BLAST database with the non-target genome sequences. For RUCS the entry point full was selected and the annotation of the target sequences was omitted. SpeciesPrimer was configured to run with the custom BLAST database, without a species list and the download and annotation step for the genome assemblies was omitted to provide comparable running conditions. The accessions of the input genome assemblies and the commands used can be found in Dataset S4. Primers used for a specificity check using Primer-BLAST (nr database limited to sequences from bacteria) were the two primer pairs with the best score in the results_best.tsv files (RUCS), the two best ranked primer pairs for SpeciesPrimer and the primers reported in the universal_primers.eprimer3 files (fdp).

RESULTS

Primer design

The SpeciesPrimer pipeline runs were completed in two to eight hours, excluding the time required for downloading and annotation of the genome assemblies. Depending on the number of genome assemblies, downloading and annotation of the genome assemblies took from 24 min (27) to 12 h 27 min (575). The average time for downloading and annotation of single assemblies was two seconds and one minute six seconds, respectively. On the consumer laptop using a wireless LAN Internet connection the time required for the downloads has doubled, while the annotation took 1.8 times longer. The pipeline runs lasted in total three times longer and were completed in seven to 29 h. The analysis of the *Enterococcus faecalis*, *Enterococcus faecium*, *Pediococcus acidilactici* and *Pediococcus pentosaceus* assemblies resulted in 15, 2, 2 and 160 identified primer pair candidates, respectively (Table 3). The primer pair candidates for *E. faecalis* and *P. pentosaceus* were filtered for the highest primer pair coverage score (*E. faecalis*: 2; *P. pentosaceus*: 29); for *P. pentosaceus*, only the two primer pairs with the lowest primer pair penalty values were selected.

The phylogeny tree from the concatenated core genes of *E. faecium* shows the phylogenetic distance of two distinct groups of sequences, a main cluster with 531 sequences and a subcluster with 44 sequences (Fig. 2). The tree made with the concatenated core gene sequences of *P. acidilactici* shows the phylogenetic distance of one sequence from all other sequences (Fig. 3). From this observation, the existence of different taxonomic units was suspected. Calculation of the average nucleotide identity (ANI) has been proposed as a valuable tool to determine species boundaries (*Richter & Rossello-Mora, 2009*). Therefore, we performed ANI calculations for the genome assemblies and displayed the results in a clustered heatmap (Fig. 4). All genome assemblies show an alignment coverage of at least 60% to each other (Dataset S3), indicating they are correctly assigned at the genus level. The clustering of the *E. faecium* genome assemblies in Fig. 4 A shows two distinct clusters corresponding to the clusters in the phylogenetic tree (Fig. 2). The assemblies of the two clusters have ANIm values at the border of the species threshold cutoff as depicted by the white to light purple colored cells. Clustering of the *P. acidilactici* genome assemblies in Fig. 4 B shows three distinct clusters corresponding to the clusters in the phylogenetic tree (Fig. 3). The purple cells indicate that the assemblies of two larger clusters belong to the same species, while the assembly with the orange color bar has ANIm percentage identity values below the proposed species threshold cutoffs (95–96%) (*Kim et al., 2014*; *Richter & Rossello-Mora, 2009*) as indicated by the blue cells. *P. acidilactici* strain FAM 18987 should therefore probably be assigned to a new species or subspecies. However, for certain species lower boundary cutoffs might be reasonable (*Ciufo et al., 2018*). According to the current taxonomic classification, we proceeded with the assumption that these genome assemblies reflected the actual diversity of strains and thus included the assemblies for the primer design.

Two test cases were generated to exemplify the influence of the input genome assemblies on the pipeline results. Firstly, a single genome assembly with a wrong taxonomic label

Table 3 Pipeline input and run statistics. Two different computers were used to run the SpeciesPrimer pipeline depicted as high end desktop and consumer laptop. The high end desktop is running Ubuntu 16.04 in a virtual machine with two Xeon E5-2643 CPU's (22 logical processors), 32 GB RAM and a solid-state drive. The download of the genome assemblies was performed using a LAN connection. The consumer laptop is running on Ubuntu 16.04 with an i7-3610QM CPU (8 logical processors), 8 GB RAM and a solid-state drive. The download of the genome assemblies was performed using a wireless LAN connection.

Species	*E. faecalis*	*E. faecium*	*P. acidilactici*	*P. pentosaceus*
Pipeline input				
NCBI genomes	390	575	9	14
ACC genomes	0	0	118	13
Total genome assemblies	**390**	**575**	**127**	**27**
Download and annotation (h:min)				
High end desktop	9:04	12:27	1:55	0:24
Consumer laptop	15:56	22:18	3:10	0:42
Pipeline statistics				
Running time (h:min)				
High end desktop	6:11	8:05	1:55	4:25
Consumer laptop	19:52	28:56	6:59	6:47
Core genes	1375	1131	921	1341
Single copy core genes	632	563	641	889
Conserved sequences	1128	624	566	2782
Species-specific sequences	329	36	54	672
Potential primer pairs	89	4	7	632
Primer pairs after QC	**15**	**2**	**2**	**160**

Notes.
QC, primer quality control; ACC, Agroscope Culture Collection.

was used as input in addition to the correctly labelled genome assemblies. Introducing a genome assembly with a wrong taxonomic label (GCF_000415325.2, *E. faecalis*) into the pool of *E. faecium* genome assemblies resulted in a decrease of identified core genes (from 1131 to 43) and provided no species-specific sequence. Secondly, the genome assembly of the *P. acidilactici* strain (FAM 18987) that was distinct from the other assemblies in the phylogenetic tree and had ANI values below the species threshold cutoff was excluded from the pipeline run. This resulted in an increased number of identified core genes (from 921 to 1238), of species-specific sequences (from 54 to 516) and of reported primer pairs (from 2 to 53).

In silico validation

Two parameters were selected as criteria for the primer validation using web-based BLAST. First, the BLAST hits for the predicted PCR product sequence should only match the target species. If sequences of other bacterial species matched to parts of the sequence, the corresponding primer pairs were discarded, unless more than three mismatches were found in each primer-binding region for the forward and reverse primers. Second, the primer binding sites in the target sequences were not allowed to have mismatches in the 3′-end region. The criterion for the primer validation by Primer-BLAST was that no predicted PCR products for other bacterial species had been reported by Primer-BLAST.

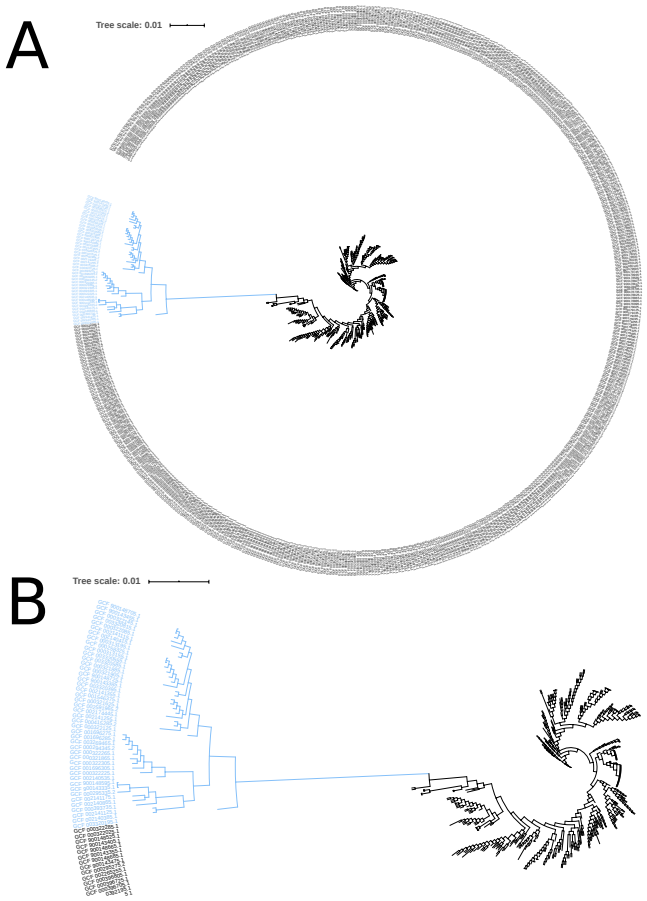


Figure 2 Phylogenetic tree based on the alignment of concatenated core genes of 575 *Enterococcus faecium* genome assemblies. (A) The main cluster with 531 sequences is depicted in black and the subcluster of 44 sequences in blue. (B) Enlarged view of the tree structure and the subcluster.
Full-size DOI: 10.7717/peerj.8544/fig-2

Primer pairs exclusively binding to the target sequence of the target species were classified as specific. The results of the *in silico* validation are summarized in Table 4. With the exception of Ec_faeca_g3060_1_P0 and Ec_faeci_cysS_3_P1, all primer pairs showed a perfect match to their target sequences. For primer pair Ec_faeca_g3060_1_P0, the first three nucleotides of one sequence out of 690 are missing in the forward primer-binding region. For Ec_faeci_cysS_3_P1, only one sequence out of 1058 aligned sequences showed a single nucleotide transition in the reverse primer-binding region (Dataset S5, page 2–3).

In vitro validation

The specificity of the qPCR assays was assessed with 21 to 25 strains of the target species (inclusivity) and 120 non-target bacterial strains found in dairy products (exclusivity). The

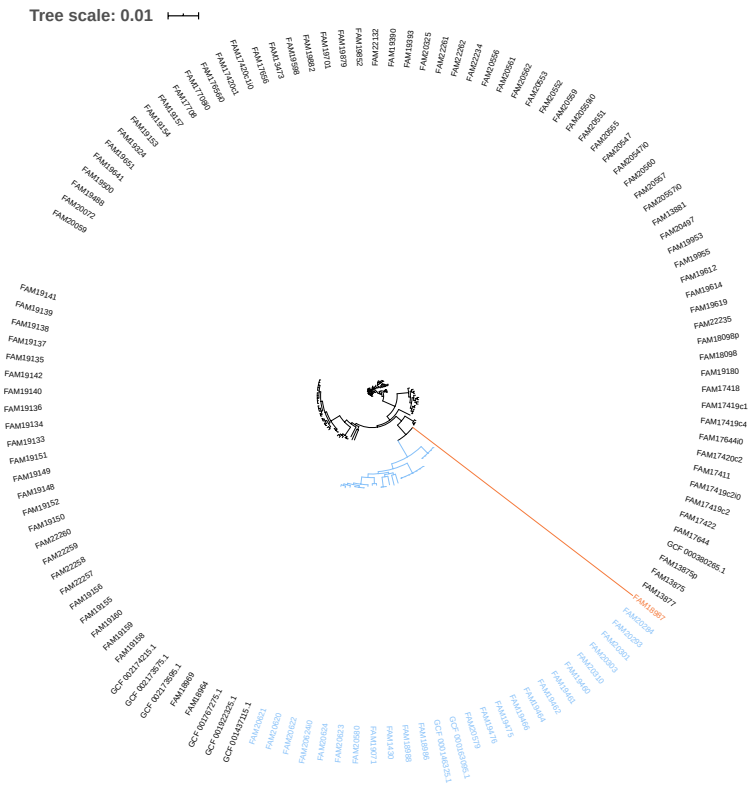


Figure 3 Phylogenetic tree based on the alignment of concatenated core genes of 127 *Pediococcus acidilactici* genome assemblies. The main cluster with 100 sequences is depicted in black, a subcluster of 26 sequences in blue and the sequence with the largest phylogenetic distance in orange.
Full-size DOI: 10.7717/peerj.8544/fig-3

qPCR assay performance was assessed by 10-fold dilution series of type strain DNA from 10^6 to 10^1 copies per reaction. The results of the qPCR runs were interpreted as positive if both qPCR reactions (duplicates) reached the fluorescent threshold before quantification cycle 35 and the peak of the melting curve analysis was above the peak calling threshold (-2 dF/dT). A summary of the results is shown in Table 5. The primer sequences can be found in Table S2. The inclusivity of the qPCR assays was 100% for the assays Ec_faeca_acuI, Ec_faeca_g3060, Ec_faeci_cysS, Pd_acidi_asnS, Pd_acidi_g1164, Pd_pento_nagK and Pd_pento_g4364. Only one qPCR assay, Ec_faeci_purD was negative for one of the tested target strains.

Out of the 120 non-target strains analyzed to determine the exclusivity of the qPCR assays (Fig. 5), all strains were negative for Ec_faeca_acuI and Pd_acidi_asnS. The assay Pd_pento_nagK targeting *P. pentosaceus* was positive for two out of three tested *Leuconostoc lactis* strains, the fluorescence signal reached the threshold after Cq 26, and the melting curve analysis showed a peak at 85 °C, while the positive control samples for this assay

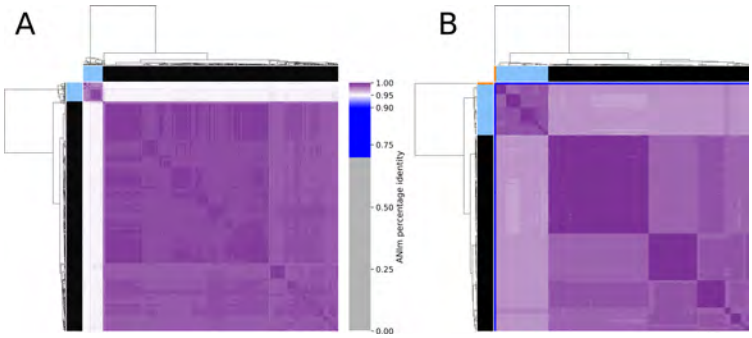


Figure 4 Clustered heatmap of ANIm percentage identity for (A) 575 *Enterococcus faecium* and (B) 129 *Pediococcus acidilactici* genome assemblies. Purple colored cells in the heatmap correspond to ANIm percentage identity above 95%, color intensity fades towards the proposed species threshold cutoff. Blue colored cells are below this threshold indicating that the corresponding genome assembly does not belong to the same species. Color bars on top and on the left of the heatmap correspond to the clusters and the colors indicated in the phylogenetic trees (Figs. 2 and 3). Dendograms are based on single-linkage hierarchical clustering of the ANIm percentage identities. Row and column names can be found in Dataset S3.

Full-size DOI: 10.7717/peerj.8544/fig-4

Table 4 Summary of the in *silico* validation of the selected primer pairs. Primer pair coverage (PPC) is a value used by MFEprimer-2.0 to score the ability of the primer pair to bind to a DNA template. The number of perfect matches of the primers to the primer binding region and the total number of target sequences are indicated in brackets.

Target species	Primer pair	PPC	BLAST (perfect/total)	Primer-BLAST (perfect/total)
E. faecalis	Ec_faeca_acuI_1_P0	100	specific (694/694)	specific (55/55)
	Ec_faeca_g3060_1_P0	100	specific (689/690)	specific (55/55)
E. faecium	Ec_faeci_cysS_3_P1	96.7	specific (1057/1058)	specific (148/148)
	Ec_faeci_purD_2_P0	93.3	specific(1083/1083)	specific (148/148)
P. acidilactici	Pd_acidi_asnS_2_P0	90.1	specific (19/19)	specific (11/11)
	Pd_acidi_g1164_1_P0	93.3	specific (23/23)	specific (11/11)
P. pentosaceus	Pd_pento_nagK_1_P0	100	specific (15/15)	specific (9/9)
	Pd_pento_g4364_1_P0	100	specific (15/15)	specific (9/9)

displayed a peak at 83.5 °C. Nine out of the 120 non-target strains were positive for the Ec_faeca_g3060 qPCR assay. For these samples the fluorescence signals reached the threshold after Cq 26 and had a melting curve peak at a higher temperature than the target PCR product. The assays Pd_acidi_g1164 and Pd_pento_g4364 were positive for five and eight non-target strains, respectively. Notably, all three tested *Lactobacillus paracasei* strains were positive for the Pd_acidi_g1164 assay, the fluorescence signal reached the threshold around Cq 21 and 22 and they showed a distinct melting curve peak at 86 °C.

The calculated efficiency of the qPCR assays was between 92 and 100%. The linear regression equations ($Cq = slope * \log(copies) + intercept$) had slopes between -3.329 and

Table 5 Summarized results of the *in vitro* validation of the selected qPCR assays. Inclusivity: Number of positive DNA samples / total number of target species DNA samples. Exclusivity: Number of DNA samples showing a fluorescence signal below quantification cycle 35 and a melting curve peak above the threshold / total number of non-target DNA samples. Calculated efficiency, slope, intercept and correlation coefficient (R^2) of the linear regression equation.

Species	*E. faecalis*		*E. faecium*		*P. acidilactici*		*P. pentosaceus*	
Target gene	*acuI*	g3060	*cysS*	*purD*	*asnS*	g1164	*nagK*	g4364
Inclusivity	22/22	22/22	25/25	24/25	21/21	21/21	25/25	25/25
Exclusivity	0/120	9/120	0/120	0/120	0/120	5/120	2/120	8/120
Efficiency	98%	97%	92%	97%	99%	100%	94%	92%
Slope	−3.382	−3.387	−3.539	−3.396	−3.356	−3.329	−3.470	−3.523
Intercept	32.107	32.694	32.006	31.051	30.835	30.282	32.286	33.211
R^2	0.998	0.997	0.990	0.996	0.997	0.995	0.996	0.997

-3.523 and correlation coefficients of 0.990 or above. Dataset S6 contains the qPCR raw data and Dataset S7 a summary of the qPCR data.

Comparison of primer design pipelines

The running times and the number of reported primer pairs of RUCS, the fdp pipeline, and SpeciesPrimer were compared. The download and annotation times were not considered since RUCS and fdp do not include this feature. RUCS and fdp were both able to design primer pairs for all four bacterial targets. The runs with RUCS were completed in two hours and 11 min to five hours and 20 min and between 107 and 629 primer candidates were reported. The specificity check using online Primer-BLAST showed that the best-ranked primer pair for each of the targets was specific and perfectly matched to the primer binding region for all targets in the nr database. The fdp runs were completed in two min to 17 h 44 min. Three primer pairs were reported for *E. faecium* and *P. acidilactici* and six primer pairs were reported for *E. faecalis* and *P. pentosaceus*. Primer-BLAST results indicate that the best primer pairs for all target species are specific. The best primer pair for *P. acidilactici* showed a two-nucleotide mismatch in the primer binding region of one target sequence. A one-nucleotide mismatch in the primer binding region of one target sequence was also observed in the primer pair for *E. faecalis* (Dataset S4, Primer-BLAST summary). The results of the SpeciesPrimer runs differ from the runs with the nt BLAST database presented in detail above. For the Enterococci the best reported primer pairs remain the same, while different primer pairs were ranked best for the Pediococci.

In summary, all pipelines were able to design species-specific primers for all of the four target species using the given input sequences. The results of the comparison are summarized in Table 6.

DISCUSSION

After setup of the SpeciesPrimer Docker container, the download of the local BLAST database and the selection of the SpeciesPrimer run settings, no further manual handling was required to get primer pair candidates for all four bacterial species after a total time of 44 h and 30 min (high end desktop). The number of input genomes and subsequently the

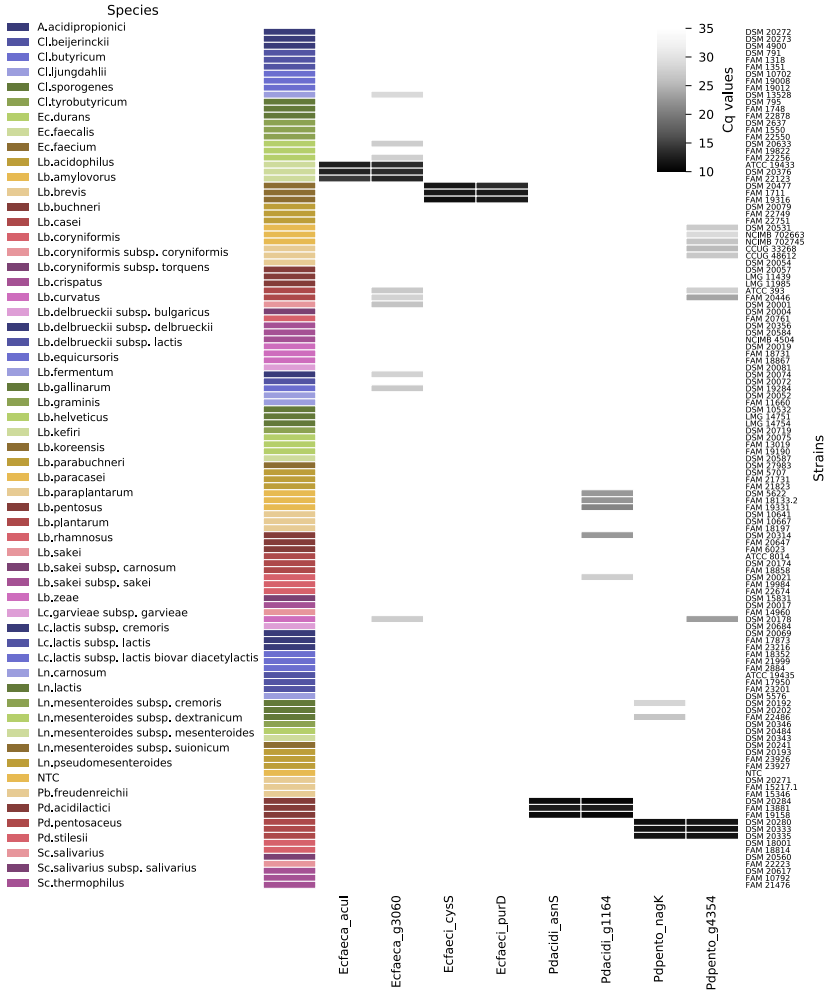


Figure 5 **qPCR assay quantification cycle heatmap.** Depicted are all tested non-target strains and their average quantification cycle (technical duplicates) if the melt curve peak was above the threshold. The gray shades represent the Cq values from 10 to 35 (if no fluorescent signal was measured the value was set to Cq 35). Abbreviations: A., *Acidipropionibacterium*; Cl., *Clostridium*; Lb., *Lactobacillus*; Ln., *Leuconostoc*; Pb, *Propionibacterium*; Pd., *Pediococcus*; Sc., *Streptococcus*; NTC, no template control.
Full-size DOI: 10.7717/peerj.8544/fig-5

number of retrieved primer pairs for the specificity check have the highest impact on speed. During the specificity check, blasting the primer sequences optimized for short sequences (blastn-short) and the subsequent compilation and indexing of the non-target sequence database are the most time consuming steps.

Table 6 Comparison of different primer design pipelines. The hardware used to run the pipelines was a high end desktop running Ubuntu 16.04 in a virtual machine with two Xeon E5-2643 CPU's (22 logical processors), 32 GB RAM and a solid-state drive. For the Primer-BLAST results, the number of perfect matches of the primers to the primer binding region and the total number of target sequences are indicated in brackets.

Target species	*Enterococcus faecalis*	*Enterococcus faecium*	*Pediococcus acidilactici*	*Pediococcus pentosaceus*
Running time (h:min)				
RUCS	5:17	5:20	2:11	3:16
fdp	8:32	17:44	0:36	0:02
SpeciesPrimer	5:23	6:53	1:35	1:11
Primer pairs				
RUCS	199	107	123	629
fdp	6	3	3	6
SpeciesPrimer	11	3	2	36
Primer-BLAST (perfect / total) Best primer pair				
RUCS	specific (59/59)	specific (153/153)	specific (11/11)	specific (9/9)
fdp	specific (58/59)	specific (153/153)	specific (10/11)	specific (9/9)
SpeciesPrimer	specific (59/59)	specific (153/153)	specific (11/11)	specific (9/9)

Notes.
fdp, find_differential_primers.

The results of the SpeciesPrimer pipeline for the four target species ranged from two to 160 identified primer pair candidates. Several factors can influence the number of identified primer pairs, such as the quality of the input genome assemblies, assemblies with wrong taxonomic labels and the genetic diversity within the species. A low-quality assembly with missing sequences or contaminations can decrease the number of identified core genes. The initial quality control helps to minimize the risk that such assemblies are included in the pipeline runs. However, also an increased sequence diversity, either due to sequencing errors, assembly errors or real diversity, limits the number and the length of identified conserved sequences. Subsequently this affects the yield of reported primer pairs, since the pipeline selects highly conserved sequences for primer design. The two test cases designed to exemplify the influence of the input genome assemblies on the pipeline results illustrate that the SpeciesPrimer pipeline performs best on closely related (same species) genome assemblies with a good overall quality.

The specificity of the designed primers was evaluated *in silico* by BLAST with a more extensive database (RefSeq Genome) than the one used for the specificity check during primer design. The validation showed that the specificity of the tested amplicons was high and no other species than the target species had an identical sequence. Most target sequences in the database showed a perfect match for the primers in the primer-binding region. For all tested primer pairs, only the expected PCR products for the target species and no amplicons for other sequences were predicted by Primer-BLAST. The results of Primer-BLAST indicate that the reported primer pairs were very specific, even though the species list used for the specificity evaluation during primer design covered only 259 non-target species.

In this work, 21 to 25 target strains for each target species and 120 non-target strains have been tested to assess inclusivity and exclusivity of the qPCR assays, respectively. The *in*

vitro validation of primer pairs has shown that the *in silico* validation is not always able to predict non-target PCR products. The fluorescence signals occurring at late quantification cycles (Cq >30) are probably due to PCR products with suboptimal primer binding. Testing the qPCR assays in mixtures and communities could be interesting to assess if these PCR products also accumulate in presence of target DNA. The specificity could be sufficient in mixtures due to competition for the primers and the difference in primer binding and amplification efficiency. For many research applications, qPCR assays with a low signal in negative samples are acceptable, assuming that low-level signals can be distinguished from low concentrations of target species DNA by the melting curve analysis (*Ririe, Rasmussen & Wittwer, 1997*). Furthermore, for many applications, the annealing temperature can be optimized by empirical determination of a suitable annealing temperature and the primer concentration can be adjusted to improve the specificity of the assay (*Bio-Rad Laboratories, 2019*). We did not try to optimize our assays with these measures, because the aim was to design primers for high-throughput qPCR, requiring the exact same PCR conditions. For the tested qPCR conditions, the most specific qPCR assays were Ec_faeca_acuI (*E. faecalis*), Ec_faeci_cysS (*E. faecium*), Pd_acidi_asnS (*P. acidilactici*) and Pd_pento_nagK (*P. pentosaceus*). Further work will be necessary in order to make these qPCR assays fully operational for the quantification of bacteria in a complex system such as food. For instance, suitable qPCR standards should be designed and validated, so that the limit of detection of each assay can be determined (*Forootan et al., 2017*).

Primer-BLAST, fdp and RUCS allow designing primers for different applications, but demand experience and manual manipulations. Primer-BLAST designs primers and performs specificity checks, but requires a target sequence provided by the user. In the case of RUCS, manual manipulation and some experience is needed to prepare the positive and negative reference sets. The same applies to fdp and the results from the comparison indicate that fdp has its strength in the identification of strain-specific primer pairs and for subsets of the positive set as implied in the name. The observed mismatches in the primer-binding region are probably due to the alignment-free approach the pipeline uses. This seems to drastically increase the speed, but it is not taking into account the conservation of the target sequence and therefore mismatches, e.g., due to single nucleotide polymorphisms (SNPs), can be found in the primer binding region. For a large number of input assemblies, e.g., for *E. faecalis* (575), fdp requires distinctively more time to run, which is a known issue caused by the cross-validation prediction step using PrimerSearch (*Pritchard et al., 2012*).

Compared to primer-BLAST and RUCS, the task SpeciesPrimer performs is really specialized, namely to design primers for species-specific sequences. In contrast, SpeciesPrimer requires no previous knowledge about the input genome assemblies and no manual manipulation of sequences has to be performed. The ability of SpeciesPrimer to run on standard computers with good performance instead of specialized high-performance computers should allow primer design for the wider range of scientists. Docker containers simplify the installation procedure and should allow non-bioinformaticians to setup and use the SpeciesPrimer pipeline.

CONCLUSIONS

In this work, we presented the SpeciesPrimer pipeline, which is a fully automated pipeline from the download of bacterial genomes, the identification of conserved species-specific core genes to primer design and subsequent quality control of primer candidates. Primers for four bacterial species were designed and validated and have shown to perform adequately under the same qPCR conditions.

A standard computer with good performance, good quality genome assemblies, a local copy of the nt BLAST database and a list of non-target bacterial species are the only requirements for primer design with SpeciesPrimer. A complete image of a Linux OS with all dependencies and the pipeline scripts is available from Dockerhub. To simplify primer design for users not familiar with command line tools, a graphic user interface is provided in the latest version of SpeciesPrimer. SpeciesPrimer facilitates efficient primer design for species-specific quantification, paving the way for a fast and accurate quantitative investigation of microbial communities.

ACKNOWLEDGEMENTS

We would like to thank Marco Meola and Remo Schmidt for critically reading the manuscript and many helpful comments and Daniel Marzohl, Nadine Sidler, Elvira Wagner and Kotchanoot Srikham for their valuable technical help.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Matthias Dreier conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Hélène Berthoud, Noam Shani, Daniel Wechsler and Pilar Junier conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Code is available at GitHub:
https://github.com/biologger/speciesprimer
Docker image is available at DockerHub:
https://hub.docker.com/r/biologger/speciesprimer.
WGS data is available at NCBI BioProject: PRJNA576774.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.8544#supplemental-information.

REFERENCES

Almeida M, Hébert A, Abraham A-L, Rasmussen S, Monnet C, Pons N, Delbès C, Loux V, Batto J-M, Leonard P, Kennedy S, Ehrlich SD, Pop M, Montel M-C, Irlinger F, Renault P. 2014. Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis of dairy fermented products. *BMC Genomics* **15**:1101 DOI 10.1186/1471-2164-15-1101.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410 DOI 10.1016/s0022-2836(05)80360-2.

Bio-Rad Laboratories. 2019. qPCR assay design and optimization. *Available at https://www.bio-rad.com/en-ch/applications-technologies/qpcr-assay-design-optimization* (accessed on 13 February 2019).

Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology* **68**:2386–2392 DOI 10.1099/ijsem.0.002809.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, De Hoon MJ. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**:1422–1423 DOI 10.1093/bioinformatics/btp163.

Cremonesi P, Pisani LF, Lecchi C, Ceciliani F, Martino P, Bonastre AS, Karus A, Balzaretti C, Castiglioni B. 2014. Development of 23 individual Taq-Man(R) real-time PCR assays for identifying common foodborne pathogens using a single set of amplification conditions. *Food Microbiology* **43**:35–40 DOI 10.1016/j.fm.2014.04.007.

Curran T, Coyle PV, McManus TE, Kidney J, Coulter WA. 2007. Evaluation of real-time PCR for the detection and quantification of bacteria in chronic obstructive pulmonary disease. *FEMS Immunology and Medical Microbiology* **50**:112–118 DOI 10.1111/j.1574-695X.2007.00241.x.

Elbrecht V, Leese F, Bunce M. 2017. PrimerMiner: an R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution* **8**:622–626 DOI 10.1111/2041-210x.12687.

Falentin H, Postollec F, Parayre S, Henaff N, Le Bivic P, Richoux R, Thierry A, Sohier D. 2010. Specific metabolic activity of ripening bacteria quantified by real-time reverse transcription PCR throughout Emmental cheese manufacture. *International Journal of Food Microbiology* **144**:10–19 DOI 10.1016/j.ijfoodmicro.2010.06.003.

Forootan A, Sjoback R, Bjorkman J, Sjogreen B, Linz L, Kubista M. 2017. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomolecular Detection and Quantification* **12**:1–6 DOI 10.1016/j.bdq.2017.04.001.

Garrido-Maestu A, Azinheiro S, Carvalho J, Prado M. 2018. Rapid and sensitive detection of viable Listeria monocytogenes in food products by a filtration-based protocol and qPCR. *Food Microbiology* **73**:254–263 DOI 10.1016/j.fm.2018.02.004.

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

Ishii S, Segawa T, Okabe S. 2013. Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. *Applied and Environmental Microbiology* **79**:2891–2898 DOI 10.1128/aem.00205-13.

Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **64**:346–351 DOI 10.1099/ijs.0.059774-0.

Kleyer H, Tecon R, Or D. 2017. Resolving species level changes in a representative soil bacterial community using microfluidic quantitative PCR. *Frontiers in Microbiology* **8**:Article 2017 DOI 10.3389/fmicb.2017.02017.

Letunic I, Bork P. 2019. Interactive tree Of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**:W256–W259 DOI 10.1093/nar/gkz239.

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, ed. *Multiple sequence alignment methods*. Totowa: Humana Press, 155–170.

Masco L, Vanhoutte T, Temmerman R, Swings J, Huys G. 2007. Evaluation of real-time PCR targeting the 16S rRNA and recA genes for the enumeration of bifidobacteria in probiotic products. *International Journal of Food Microbiology* **113**:351–357 DOI 10.1016/j.ijfoodmicro.2006.07.021.

Moyaert H, Pasmans F, Ducatelle R, Haesebrouck F, Baele M. 2008. Evaluation of 16S rRNA gene-based PCR assays for genus-level identification of helicobacter species. *Journal of Clinical Microbiology* **46**:1867–1869 DOI 10.1128/jcm.00139-08.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**:3691–3693 DOI 10.1093/bioinformatics/btv421.

Postollec F, Falentin H, Pavan S, Combrisson J, Sohier D. 2011. Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiology* **28**:848–861 DOI 10.1016/j.fm.2011.02.008.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**:e9490 DOI 10.1371/journal.pone.0009490.

Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8**:12–24 DOI 10.1039/c5ay02550h.

Pritchard L, Holden NJ, Bielaszewska M, Karch H, Toth IK. 2012. Alignment-free design of highly discriminatory diagnostic primer sets for Escherichia coli O104:H4 outbreak strains. *PLOS ONE* **7**:e34498 DOI 10.1371/journal.pone.0034498.

Qu W, Zhou Y, Zhang Y, Lu Y, Wang X, Zhao D, Yang Y, Zhang C. 2012. MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Research* **40**:W205–W208 DOI 10.1093/nar/gks552.

Ramirez M, Castro C, Palomares JC, Torres MJ, Aller AI, Ruiz M, Aznar J, Martin-Mazuelos E. 2009. Molecular detection and identification of Aspergillus spp. from clinical samples using real-time PCR. *Mycoses* **52**:129–134 DOI 10.1111/j.1439-0507.2008.01548.x.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* **16**:276–277 DOI 10.1016/S0168-9525(00)02024-2.

Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**:19126–19131 DOI 10.1073/pnas.0906412106.

Ririe KM, Rasmussen RP, Wittwer CT. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry* **245**:154–160 DOI 10.1006/abio.1996.9916.

Sayers E. 2009. The E-utilities in-depth: parameters, syntax and more. *Available at https://www.ncbi.nlm.nih.gov/books/NBK25499/*.

Scheirlinck I, Van der Meulen R, De Vuyst L, Vandamme P, Huys G. 2009. Molecular source tracking of predominant lactic acid bacteria in traditional Belgian sourdoughs and their production environments. *Journal of Applied Microbiology* **106**:1081–1092 DOI 10.1111/j.1365-2672.2008.04094.x.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**:2068–2069 DOI 10.1093/bioinformatics/btu153.

Shen Z, Qu W, Wang W, Lu Y, Wu Y, Li Z, Hang X, Wang X, Zhao D, Zhang C. 2010. MPprimer: a program for reliable multiplex PCR primer design. *BMC Bioinformatics* **11**:143 DOI 10.1186/1471-2105-11-143.

Tange O. 2011. GNU parallel—the command-line power tool. *login: the USENIX Magazine* **36**:42–47.

Tao T, Madden T, Christiam C, Szilagyi L. 2011. BLAST FTP site. *Available at https://www.ncbi.nlm.nih.gov/books/NBK62345/* (accessed on 24 September 2019).

Thomsen MCF, Hasman H, Westh H, Kaya H, Lund O. 2017. RUCS: rapid identification of PCR primers for unique core sequences. *Bioinformatics* **33**:3917–3921 DOI 10.1093/bioinformatics/btx526.

Torriani S, Felis GE, Dellaglio F. 2001. Differentiation of Lactobacillus plantarum, L. pentosus, and L. paraplantarum by recA gene sequence analysis and multiplex PCR

assay with recA gene-derived primers. *Applied and Environmental Microbiology* **67**:3450–3454 DOI 10.1128/aem.67.8.3450-3454.2001.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**:e115 DOI 10.1093/nar/gks596.

Vijaya Satya R, Kumar K, Zavaljevski N, Reifman J. 2010. A high-throughput pipeline for the design of real-time PCR signatures. *BMC Bioinformatics* **11**:340 DOI 10.1186/1471-2105-11-340.

Wang LT, Lee FL, Tai CJ, Kasai H. 2007. Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group. *International Journal of Systematic and Evolutionary Microbiology* **57**:1846–1850 DOI 10.1099/ijs.0.64685-0.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**:134 DOI 10.1186/1471-2105-13-134.

Zhu T, Liang C, Meng Z, Li Y, Wu Y, Guo S, Zhang R. 2017. PrimerServer: a high-throughput primer design and specificity-checking platform. bioRxiv:181941 DOI 10.1101/181941.

Zuker M, Mathews DH, Turner DH. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark BFC, eds. *RNA biochemistry and biotechnology*. Dordrecht: Springer Netherlands, 11–43.

2.9 Supplementary Information

Supplementary data set S1

SpeciesPrimer input and parameters

The data set S1 is a large Excel workbook containing multiple worksheets with tabular data, it is available at https://doi.org/10.7717/peerj.8544/supp-1. The data set is too large to include it in the thesis. A short description of the content is given below.

Sheets:

1. NCBI accessions and FAM numbers of the input genome assemblies for *Enteroccus faecium*, *Enterococcus faecalis*, *Pediococcus acidilactici* and *Pediococcus pentosaceus* used for primer design with SpeciesPrimer.
2. NCBI WGS accessions for FAM strains uploaded to NCBI
3. A summary of the SpeciesPrimer pipeline settings used for primer designed
4. Species list for species used for the *in silico* specificity checks by the SpeciesPrimer pipeline

Supplementary data set S2

Bacterial strains used in the study

The data set S2 is a large Excel workbook containing multiple worksheets with tabular data, it is available at https://doi.org/10.7717/peerj.8544/supp-2. The data set is too large to include it in the thesis. A short description of the content is given below.

Sheets:

1. List of target strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.
2. List of non-target strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.
3. Cultivation conditions for the strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.

Supplementary data set S3

Average nucleotide identity data and newick trees

The data set S3 is a ZIP archive containing multiple files with different types of data, it is available at https://doi.org/10.7717/peerj.8544/supp-3. The data set is too large to include it in the thesis. A short description of the content is given below.

1. ANIm_data
   - Alignment coverage tables of ANI analysis for *E. faecium* and *P. acidilactici*
     - Ecfaeci_ANIm_alignment_coverage.tab
     - Pdacidi_ANIm_alignment_coverage.tab
   - Labels for the ANI analysis heatmaps in Figure 4 for *E. faecium* and *P. acidilactici*
     - Ecfaeci_ANIm_labels_Fig4.csv
     - Pdacidi_ANIm_labels_Fig4.csv
   - ANIm percentage identity tables for *E. faecium* and *P. acidilactici* (raw data for Figure 4.)
     - Ecfaeci_ANIm_percent_identity.tab
     - Pdacidi_ANIm_percent_identity.tab
2. Newick_trees
   - Phylogenetic trees used for Figure 2 and Figure 3 in newick format
     - Ec_faecium_tree.nwk
     - Pd_acidilactici_tree.nwk

Supplementary data set S4

Pipeline comparison input and parameters

The data set S4 is a large Excel workbook containing multiple worksheets with tabular data, it is available at https://doi.org/10.7717/peerj.8544/supp-4. The data set is too large to include it in the thesis. A short description of the content is given below.

Sheets:

1. NCBI accessions and FAM numbers of the input genome assemblies for *Enteroccus faecium*, *Enterococcus faecalis*, *Pediococcus acidilactici* and *Pediococcus pentosaceus* used for primer design with RUCS and find_differential_primers.
2. NCBI accessions of the negative set genome assemblies used for the evaluation of primer specificity by RUCS and find_differential_primers.
3. Software dependencies used to run RUCS and find_differential_primers.
4. Commands used to run SpeciesPrimer, RUCS and find_differential_primers.
5. Primers with the highest scores reported by SpeciesPrimer, RUCS and find_differential_primers.
6. Summary of the primer-BLAST results for the primers with the highest scores reported by SpeciesPrimer, RUCS and find_differential_primers.

Supplementary data set S5

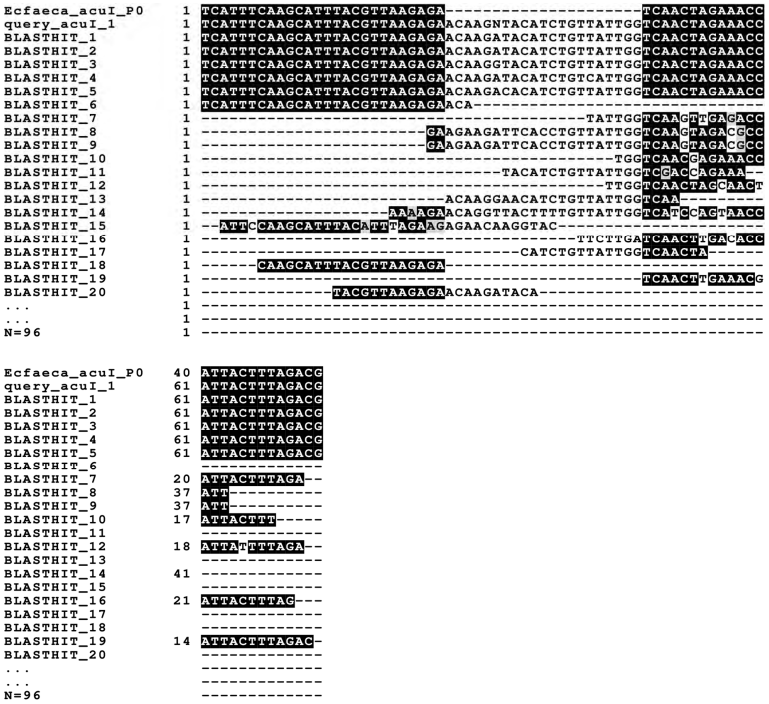
Summary of predicted-amplicon BLAST results.

Summarized alignments of web-based BLAST search against the RefSeq Genomes database (limited to Bacteria, taxid:2) with the predicted PCR product of the primer pairs as query. The first line in the alignment represent the forward primer sequence (left) and the reverse-complement sequence of the reverse primer (right). The second line represents the predicted PCR product. Only one sequence per BLAST hit for a certain species is shown if the sequences were identical, the number of these redundant hits is given in the last row of the results table. The file is also available at https://doi.org/10.7717/peerj.8544/supp-5.

Summarized results of predicted amplicon BLAST search against the RefSeq Genomes database (limited to Bacteria, taxid:2).

Summarized alignments of web-based BLAST search with the predicted PCR product of the primer pairs as query. The first line in the alignment represent the forward primer sequence (left) and the reverse-complement sequence of the reverse primer (right). The second line represents the predicted PCR product. Only one sequence per BLAST hit for a certain species is shown if the sequences were identical, the number of these redundant hits is given in the last row of the results table.

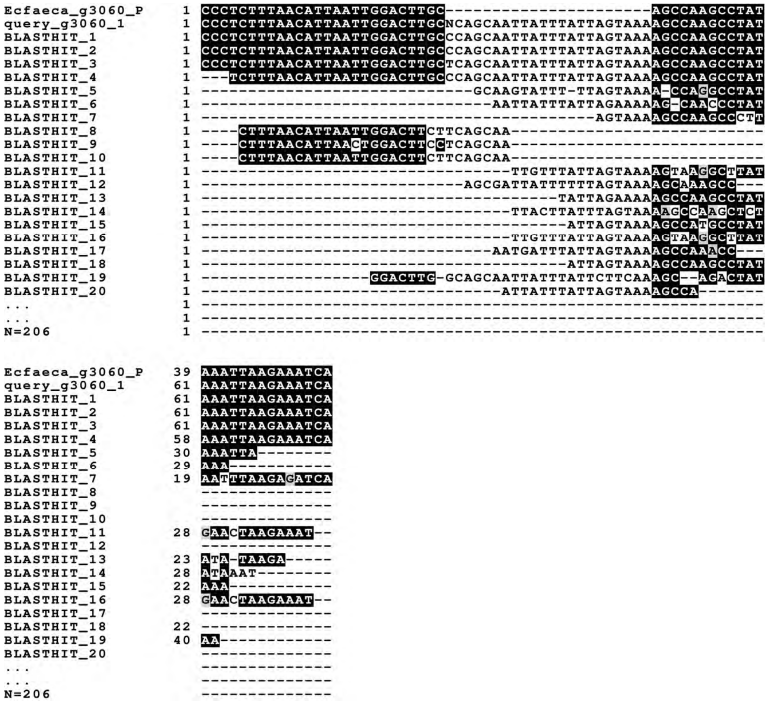
Ec_faeca_acuI_1_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Enterococcus faecalis	100	99	NC_004668.1	690 x
2	Enterococcus sp.	100	99	NZ_JWBQ01000067.1	3 x
3	Enterococcus faecalis	100	99	NZ_JUXC01000095.1	-
4	Enterococcus faecalis	100	97	NZ_PHLF01000001.1	-
5	Enterococcus faecalis	100	97	NZ_KZ845811.1	-
6	Staphylococcus aureus	40	100	NZ_LJOC01000115.1	1 x
7	Anaerorhabdus furcosa	41	90	NZ_FUWY01000002.1	-
8	Oceanihabitans sediminis	53	82	NZ_QPIG01000005.1	1 x
9	Flavobacteriaceae bacterium	53	82	NZ_MPGM01000017.1	-
10	Nitrosotalea sp.	33	96	NZ_FRFC01000001.1	-
11	Collimonas fungivorans	36	92	NZ_CP013232.1	-
12	Nostoc piscinale	38	89	NZ_CP012036.1	-
13	Ewingella americana	34	92	NZ_JMPJ01000065.1	-
14	Neisseriaceae bacterium	55	80	NZ_CP024847.1	-
15	Clostridium kluyveri	46	80	NZ_CP018335.1	-
16	Planktothrix agardhii	41	87	NZ_CM002803.1	-
17	Hungateiclostridium cellulolyticum	27	100	NZ_JH556653.1	-
18	Pseudoalteromonas atlantica	27	100	NC_008228.1	-
19	Bacillus simplex	34	92	NZ_NQLS01000166.1	2 x
20	Olleya namhaensis	30	95	NZ_FORM01000001.1	-

* anomalous sequences excluded from alignment
gi|1423449487|ref|NZ_UEMU01000447.1|Escherichia coli strain KCRI-77E isolate RDK43_77E
gi|1005608163|ref|NZ_FFHQ01000001.1|Listeria monocytogenes strain 2842STDY5753961
gi|522836524|ref|NZ_KE352150.1|Enterococcus faecalis strain SB2C-2 Scaffold11
gi|1120478950|ref|NZ_FQTZ01000002.1|Enterococcus faecium isolate Hp_76-17_S13_

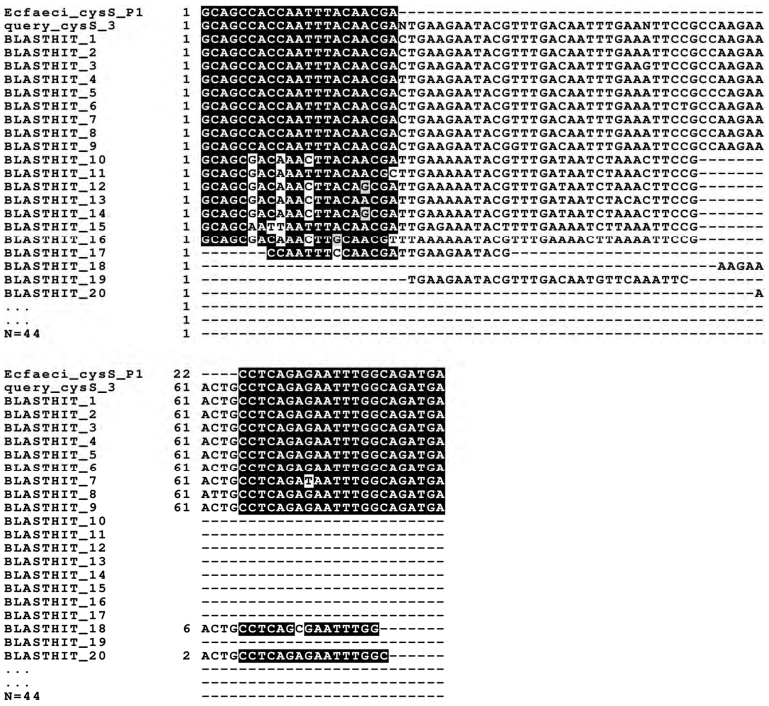
Ec_faeca_g3060_1_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Enterococcus faecalis	100	99	NC_004668.1	687 x
2	Enterococcus sp.	100	99	NZ_JWBQ01000066.1	2 x
3	Enterococcus faecalis	100	99	NZ_GL454749.1	-
4	Enterococcus faecalis	96	99	NZ_PTYL01000009.1	-
5	Flavobacterium sp.	48	87	NZ_CP031557.1	-
6	Buchnera aphidicola	42	90	NC_004061.1	1 x
7	Acinetobacter johnsonii	43	88	NZ_FUUY01000018.1	-
8	Pasteurella multocida subsp. septica	39	90	NZ_UGSW01000001.1	1 x
9	Pasteurella multocida	39	90	NZ_PSQH01000001.1	4 x
10	Pasteurella multocida	39	90	NZ_CP023972.1	-
11	Enterococcus thailandicus	53	82	NZ_LWMN01000016.1	2 x
12	Algoriphagus halophilus	39	90	NZ_FSRC01000001.1	-
13	Candidatus Borrelia	41	89	NZ_CP025785.1	-
14	Elizabethkingia ursingii	43	89	NZ_MBDS01000002.1	-
15	Clostridiales bacterium	32	96	NZ_JQKL01000001.1	-
16	Enterococcus sp.	53	82	NZ_NGMR01000001.1	-
17	Flavobacterium beibuense	35	92	NZ_JRLV01000001.1	-
18	Aliterella atlantica	28	100	NZ_JYON01000010.1	-
19	Herpetosiphon geysericola	56	79	NZ_LGKP01000017.1	-
20	Corynebacterium kutscheri	28	100	NZ_CP011312.1	-

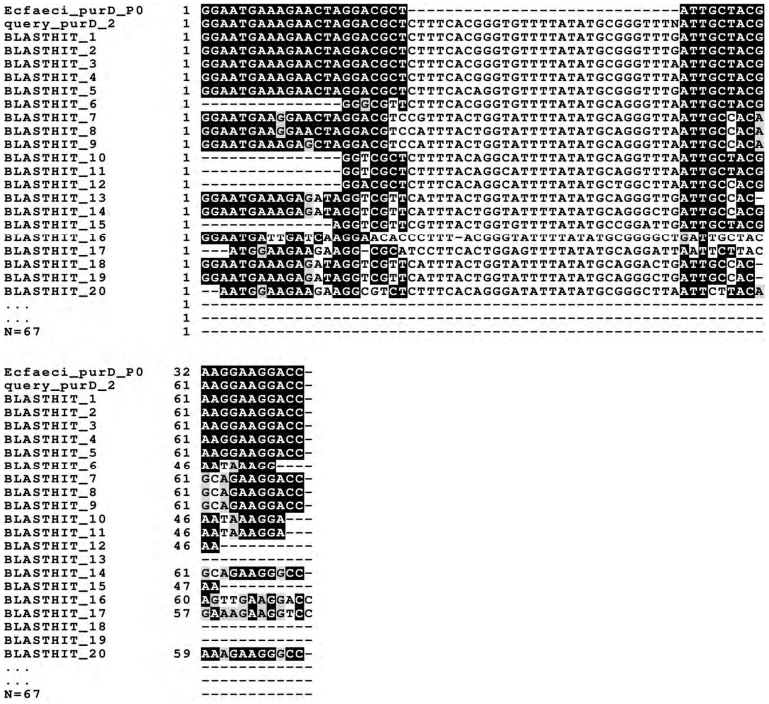
* anomalous sequences excluded from alignment
gi|1005608163|ref|NZ_FFHQ01000001.1|Listeria monocytogenes strain 2842STDY5753961
gi|522836690|ref|NZ_KE352316.1|Enterococcus faecalis strain SB2C-2 Scaffold89
gi|1120479204|ref|NZ_FQTZ01000012.1|Enterococcus faecium isolate Hp_76-17_S13
gi|1129541613|ref|NZ_CWML01000029.1|Listeria monocytogenes isolate LM52

Ec_faeci_cysS_3_P1



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Enterococcus faecium	100	98	NC_017960.1	1051 x
2	Enterococcus sp.	100	98	NZ_AJRB03000003.1	37 x
3	Enterococcus faecium	100	98	NZ_PTVC01000111.1	-
4	Enterococcus faecium	100	98	NZ_KB029683.1	-
5	Enterococcus sp.	100	97	NZ_KV829502.1	-
6	Enterococcus faecium	100	97	NZ_PUAH01000015.1	-
7	Enterococcus faecium	100	97	NZ_PUBG01000061.1	-
8	Enterococcus faecium	100	97	NZ_PTTS01000223.1	-
9	Enterococcus faecium	100	97	NZ_MJDZ01000007.1	-
10	Enterococcus mundtii	62	83	NZ_PTUS01000014.1	17 x
11	Enterococcus mundtii	62	83	NZ_PYGS01000082.1	-
12	Enterococcus mundtii	62	81	NC_022878.1	-
13	Enterococcus mundtii	62	81	NZ_AFWZ01000346.1	-
14	Enterococcus pernyi	62	81	NZ_KV388085.1	-
15	Enterococcus ratti	62	79	NZ_JXLB01000002.1	-
16	Enterococcus villorum	62	77	NZ_MJEB01000016.1	1 x
17	Bacillus okhensis	30	92	NZ_JRJU01000050.1	-
18	Anaplasma phagocytophilum	28	96	NZ_LANV01000001.1	17 x
19	Chitinophaga terrae	34	89	NZ_FNRL01000002.1	-
20	Amycolatopsis sacchari	24	100	NZ_FORP01000012.1	-

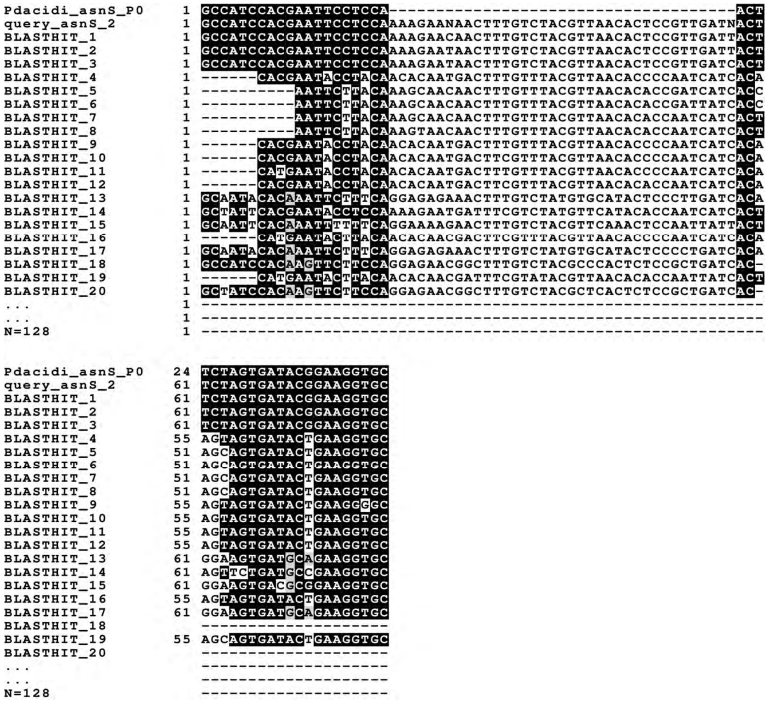
Ec_faeci_purD_2_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Enterococcus faecium	100	99	NC_017960.1	1080 x
2	Enterococcus sp.	100	99	NZ_AJRB03000044.1	37 x
3	Enterococcus sp.	100	99	NZ_NGMF01000001.1	-
4	Enterococcus faecium	100	99	NZ_JH804745.1	-
5	Enterococcus faecium	100	97	NZ_JWEB01000121.1	-
6	Bacillus cihuensis	75	87	NZ_AYSD01000010.1	-
7	Enterococcus mundtii	100	79	NZ_PYGS01000033.1	5 x
8	Enterococcus mundtii	100	79	NZ_FOUC01000005.1	-
9	Enterococcus mundtii	100	79	NZ_NGMS01000001.1	-
10	Paenibacillus sp.	76	83	NZ_LMVB01000002.1	-
11	Bacillus sp.	76	83	NZ_LMXG01000002.1	-
12	Bacillus loiseleuriae	66	85	NZ_LFZW01000001.1	-
13	Enterococcus pernyi	83	80	NZ_KV388085.1	-
14	Enterococcus mundtii	100	76	NZ_PTUS01000023.1	-
15	Enterococcus faecalis	68	83	NZ_KE351686.1	-
16	Enterococcus gallinarum	99	76	NZ_BCQE01000018.1	9 x
17	Ureibacillus thermosphaericus	95	76	NZ_AJIK01000018.1	1 x
18	Enterococcus mundtii	83	78	NZ_AFWZ01000136.1	-
19	Enterococcus mundtii	83	78	NZ_PYGU01000006.1	-
20	Lysinibacillus sp.	97	75	NZ_LT985980.1	-

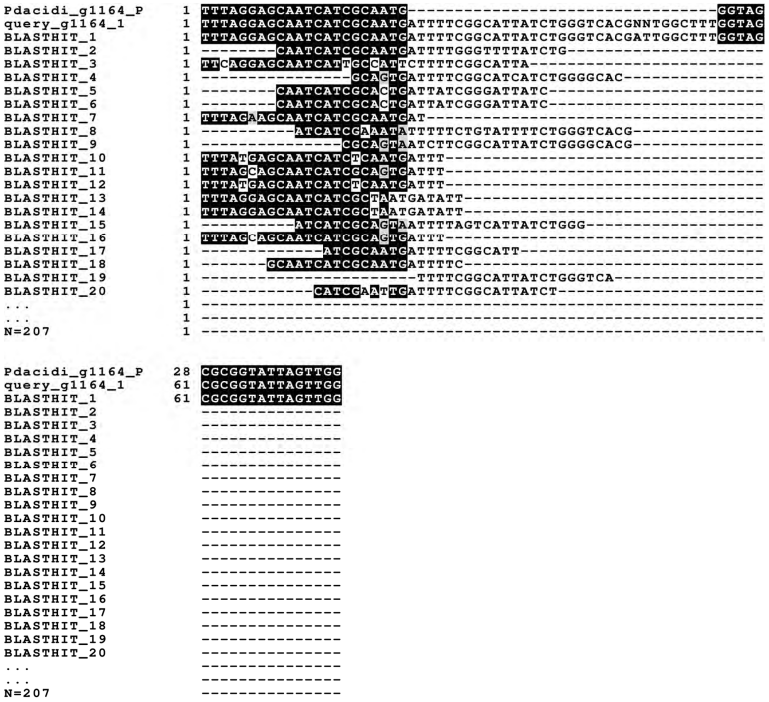
* anomalous sequences excluded from alignment
gi|1129541523|ref|NZ_CWML01000013.1|Listeria monocytogenes isolate LM52

Pd_acidi_asnS_2_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Pediococcus acidilactici	100.0	98.0	NZ_CP015206.1	16 x
2	Pediococcus acidilactici	100.0	98.0	NZ_QYRQ01000020.1	-
3	Pediococcus acidilactici	100.0	98.0	NZ_JQCC01000002.1	-
4	Lactobacillus kunkeei	92.0	78.0	NZ_CP012920.1	11 x
5	Pediococcus pentosaceus	88.0	79.0	NC_008525.1	11 x
6	Pediococcus pentosaceus	88.0	79.0	NZ_JDVW01000006.1	-
7	Pediococcus pentosaceus	88.0	79.0	NZ_JQBF01000008.1	-
8	Pediococcus pentosaceus	88.0	79.0	NC_022780.1	-
9	Lactobacillus kunkeei	92.0	77.0	NZ_JXCZ01000012.1	-
10	Lactobacillus kunkeei	92.0	77.0	NZ_JXDA01000004.1	-
11	Lactobacillus kunkeei	92.0	77.0	NZ_BDDX01000015.1	-
12	Lactobacillus kunkeei	92.0	77.0	NZ_JXDD01000007.1	-
13	Ruminiclostridium cellobioparum subsp. termitidis	100.0	74.0	NZ_AORV01000003.1	-
14	Lactobacillus sp.	100.0	74.0	NZ_CP031933.1	-
15	Leptotrichia sp.	100.0	74.0	NZ_CP016753.1	-
16	Lactobacillus kunkeei	92.0	74.0	NZ_JXDC01000005.1	-
17	Ruminiclostridium cellobioparum	100.0	72.0	NZ_JHYD01000026.1	-
18	Faecalibacterium prausnitzii	74.0	76.0	NZ_QVFB01000007.1	9 x
19	Lactobacillus sp.	92.0	73.0	NZ_CP032626.1	-
20	Faecalibacterium prausnitzii	74.0	76.0	NZ_NMTZ01000018.1	-

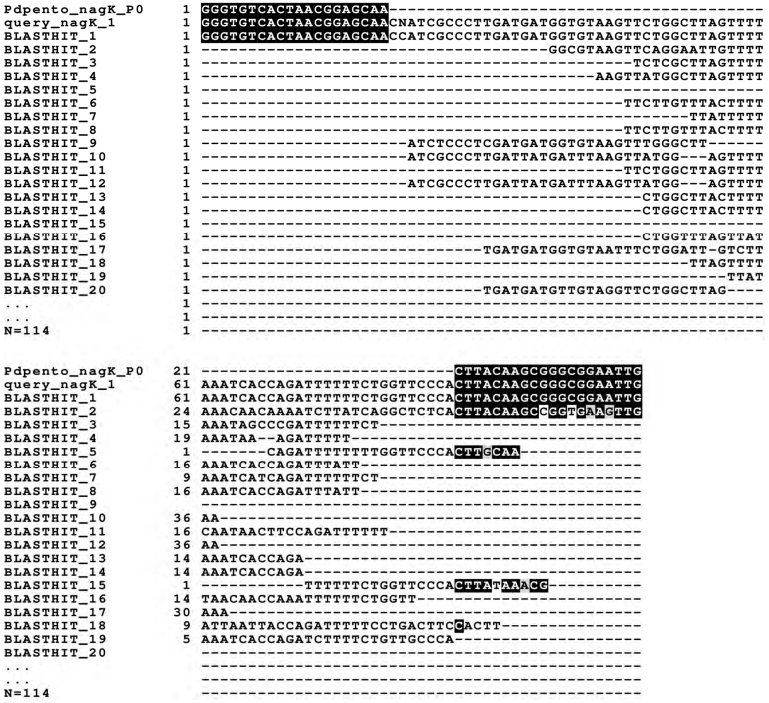
Pd_acidi_g1164_1_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Pediococcus acidilactici	100.0	97.0	NZ_CP015206.1	22 x
2	Helicobacter sp.	41.0	90.0	NZ_MLAM01000013.1	-
3	Gallibacterium genomosp.	47.0	86.0	NZ_JTJS01000032.1	-
4	Erwinia iniecta	39.0	90.0	NZ_JRXE01000016.1	1 x
5	Yersinia kristensenii	39.0	90.0	NZ_CWJK01000005.1	4 x
6	Yersinia frederiksenii	39.0	90.0	NZ_CP009364.1	1 x
7	Bacillus vietnamensis	32.0	96.0	NZ_LIXZ01000003.1	-
8	Catellicoccus marimammalium	48.0	83.0	NZ_AMYT01000022.1	-
9	Erwinia persicina	41.0	87.0	NZ_BCTN01000009.1	1 x
10	Megamonas rupellensis	35.0	92.0	NZ_KB899597.1	3 x
11	Acinetobacter kyonggiensis	35.0	92.0	NZ_FNPK01000003.1	-
12	Megamonas sp.	35.0	92.0	NZ_FOCY01000013.1	-
13	Macrococcus epidermidis	36.0	92.0	NZ_PZJH01000003.1	-
14	Macrococcus goetzii	36.0	92.0	NZ_MJBI02000002.1	-
15	Sphingobacterium faecium	41.0	87.0	NZ_QBKH01000003.1	-
16	Acinetobacter sp.	35.0	92.0	NZ_NEGM01000001.1	1 x
17	Bacillus sp.	28.0	100.0	NZ_LMTJ01000002.1	9 x
18	Pseudomonas batumici	28.0	100.0	NZ_JXDG01000015.1	-
19	Rhodovulum sp.	28.0	100.0	NZ_QPLK01000071.1	-
20	Roseovarius sp.	35.0	92.0	NZ_OUMZ01000007.1	-

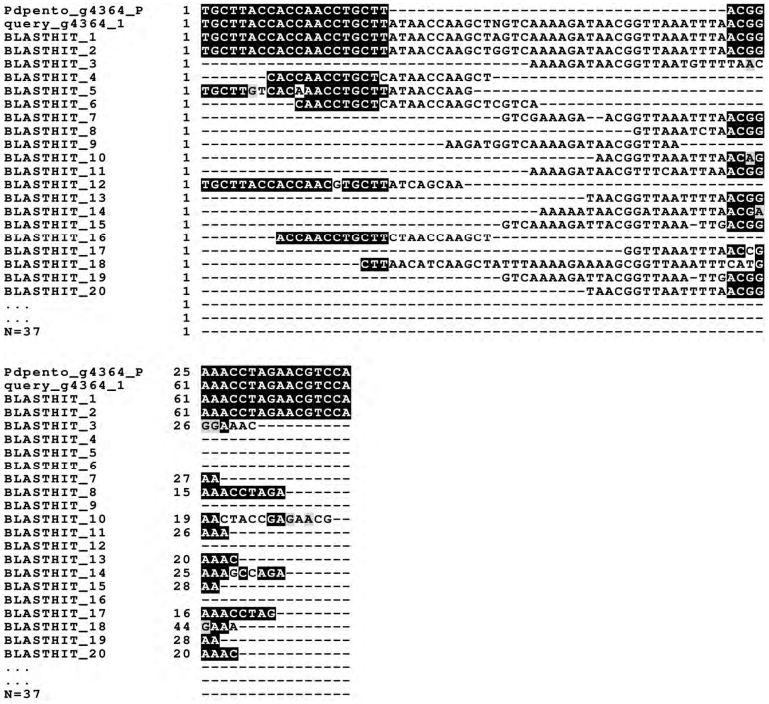
* anomalous sequences excluded from alignment
gi|522836680|ref|NZ_KE352306.1|Enterococcus faecalis strain SB2C-2 Scaffold825

Pd_pento_nagK_1_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Pediococcus pentosaceus	100.0	99.0	NC_008525.1	14 x
2	Pediococcus stilesii	65.0	74.0	NZ_JQBX01000011.1	-
3	Zhouia amylolytica	31.0	88.0	NZ_FPAG01000004.1	-
4	Clostridium perfringens	30.0	86.0	NZ_CYYX01000001.1	4 x
5	Flavobacterium haoranii	25.0	93.0	NZ_FQZH01000005.1	-
6	Acinetobacter bereziniae	30.0	88.0	NZ_QWFV01000007.1	3 x
7	Mycoplasma auris	25.0	93.0	NZ_QKUB01000015.1	-
8	Acinetobacter sp.	30.0	88.0	NZ_AMFQ01000033.1	4 x
9	Thiomonas intermedia	30.0	88.0	NC_014153.1	-
10	Leptospira meyeri	34.0	80.0	NZ_AKXE01000002.1	1 x
11	Chryseobacterium scophthalmum	30.0	83.0	NZ_FSRQ01000001.1	-
12	Leptospira sp.	34.0	80.0	NZ_NPEB01000007.1	2 x
13	Acinetobacter brisouii	22.0	96.0	NZ_JZRE01000019.1	-
14	Acinetobacter sp.	22.0	96.0	NZ_KB851219.1	-
15	Clostridium pasteurianum	24.0	92.0	NC_021182.1	-
16	Mycoplasma pullorum	34.0	83.0	NZ_CP017813.1	1 x
17	Lutibacter profundi	30.0	87.0	NZ_CP013355.1	-
18	'Nostoc azollae'	35.0	80.0	NC_014248.1	-
19	Elizabethkingia meningoseptica	29.0	87.0	NZ_BARD01000013.1	19 x
20	Sporomusa silvacetica	24.0	92.0	NZ_LSLK01000054.1	-

Pd_pento_g4364_1_P0



BLAST HIT	Species	Query coverage [%]	Identity [%]	Accession	redundant hits
1	Pediococcus pentosaceus	100.0	99.0	NC_008525.1	13 x
2	Pediococcus pentosaceus	100.0	99.0	NZ_JDVW01000004.1	-
3	Firmicutes bacterium	39.0	88.0	NZ_QUDS01000002.1	-
4	Pseudovibrio hongkongensis	32.0	96.0	NZ_LLWC01000022.1	-
5	Luteibacter rhizovicinus	38.0	90.0	NZ_JPLB01000036.1	1 x
6	Burkholderia multivorans	34.0	92.0	NZ_PVGD01000058.1	-
7	Streptococcus gallolyticus	37.0	88.0	NZ_FNFJ01000002.1	1 x
8	Unicellular cyanobacterium	30.0	96.0	NZ_NETF01000197.1	-
9	Staphylococcus gallinarum	33.0	92.0	NZ_JXCF01000010.1	1 x
10	Pyrinomonas methylaliphatogenes	40.0	86.0	NZ_CBXV010000006.1	-
11	Maribacter sedimenticola	37.0	89.0	NZ_FZNV01000002.1	-
12	Mannheimia haemolytica	37.0	89.0	NC_021082.1	66 x
13	Hungateiclostridium thermocellum	30.0	96.0	NC_009012.1	6 x
14	Virgibacillus siamensis	43.0	85.0	NZ_FUIH01000007.1	-
15	Streptococcus sanguinis	38.0	89.0	NZ_GL890988.1	-
16	Bacillus sp.	30.0	96.0	NZ_QPJG01000005.1	-
17	Bacillus taeanensis	30.0	96.0	NZ_QOCW01000025.1	-
18	Actinobacillus minor	61.0	78.0	NZ_ACQL01000172.1	-
19	Streptococcus sp.	38.0	89.0	NZ_LSJN01000035.1	-
20	Ruminiclostridium thermocellum	30.0	96.0	NZ_CP016502.1	1 x

***Anomalous sequences excluded from alignment**

gi|522836524|ref|NZ_KE352150.1|Enterococcus faecalis strain SB2C-2 Scaffold11 **gi|522836690|ref|NZ_KE352316.1|Enterococcus faecalis strain SB2C-2 Scaffold89** **gi|522836680|ref|NZ_KE352306.1|Enterococcus faecalis strain SB2C-2 Scaffold825** **Excluded because we suspected a wrong taxonomic classification.** The classification was corrected on 2018-10-26. Enterococcus faecium SB2C-2 before 2018-10-26 (see Taxonomic-Update-Statistics)
gi|1120478950|ref|NZ_FQTZ01000002.1|Enterococcus faecium isolate Hp_76-17_S13_ **gi|1120479204|ref|NZ_FQTZ01000012.1|Enterococcus faecium isolate Hp_76-17_S13_** **Excluded because we suspect a wrong taxonomic classification.** Taxonomy report of BLAST search (using default values, Nucleotide collection) of this sequence (NZ_FQTZ01000002.1) shows best hits for Enterococcus faecalis and no hits for Enterococcus faecium. BLAST search (using default values, Nucleotide collection) of this sequence (NZ_FQTZ01000012.1) shows best hits for Enterococcus faecalis and less similarity for Enterococcus faecium genome assemblies.
gi|1423449487|ref|NZ_UEMU01000447.1|Escherichia coli strain KCRI-77E isolate RDK43_77E **Excluded because we suspect a contamination.** Taxonomy report of BLAST search (using default values, Nucleotide collection) of this sequence (NZ_UEMU01000447.1) shows only results for Enterococcus faecalis.
gi|1129541523|ref|NZ_CWML01000013.1|Listeria monocytogenes isolate LM52 **gi|1129541613|ref|NZ_CWML01000029.1|Listeria monocytogenes isolate LM52** **Excluded because we suspect a wrong taxonomic classification.** The corresponding assembly (GCA_001495195.1) shows only 79.251 % gapped identity to the closest related Listeria monocytogenes assembly. BLAST search (using default values, Nucleotide collection) of these sequences (NZ_CWML01000013.1, NZ_CWML01000029.1) shows high similarity with Enterococcus thailandicus and some similarity to other Enterococcus species but no hits for Listeria species.
gi|1005608163|ref|NZ_FFHQ01000001.1|Listeria monocytogenes strain 2842STDY5753961 **Excluded because we suspect a wrong taxonomic classification.** The corresponding assembly (GCA_900017475.1) shows only 87.1677 % gapped identity to the closest related Listeria monocytogenes assembly. Taxonomy report of BLAST search (using default values, Nucleotide collection) of the first 100000 bases of this sequence (NZ_FFHQ01000001.1**)** shows best hits for Enterococcus faecalis and no hits for Listeria species.

Supplementary data set S6

Raw data of qPCR experiments

The data set S6 is a large Excel workbook containing multiple worksheets with tabular data, it is available at https://doi.org/10.7717/peerj.8544/supp-6. The data set is too large to include it in the thesis. A short description of the content is given below.

Raw qPCR data for the *in vitro* validation of the primer pairs Ec_faeca_acuI, Ec_faeca_g3060, Ec_faeci_cysS, Ec_faeci_purD, Pd_acidi_asnS, Pd_acidi_g1164, Pd_pento_nagK and Pd_pento_g4364. The quantification cycle values and melting temperatures of the melt analysis are given in different sheets for the target strains, dilution series of the target type strains and the non-target strains (divided into four parts).

Supplementary data set S7

Summary of qPCR results

The data set S7 is a large Excel workbook containing multiple worksheets with tabular data, it is available at https://doi.org/10.7717/peerj.8544/supp-7. The data set is too large to include it in the thesis. A short description of the content is given below.

An overview of the *in vitro* validation qPCR results for the primer pairs Ec_faeca_acuI, Ec_faeca_g3060, Ec_faeci_cysS, Ec_faeci_purD, Pd_acidi_asnS, Pd_acidi_g1164, Pd_pento_nagK and Pd_pento_g4364. Each sheet contains a summary of the quantification cycle values and melting temperatures for the strains of the inclusivity and exclusivity analysis, the dynamic range and efficiency of the primer pair and statistics for the run to run variation of the positive control samples.

Supplementary table S1

Table S2.1: Overview of the tools and features used for the SpeciesPrimer pipeline, available at https://doi.org/10.7717/peerj.8544/supp-8.

Tools	Features	Pipeline task
NCBI Entrez module (Biopython)	Offers a variety of tools and code to access NCBI data	Search and download of genome assemblies
BLAST+	Compares sequences and finds regions of similarity in sequence databases	Assess the similarity of input sequences to sequences in the nt database
Prokka	Fast whole genome annotation for bacterial, archeal and viral genomes	Annotation of bacterial input genome assemblies
Roary	Fast pan genome pipeline working on standard desktop PC's	Identify single copy core genes of the target species.
FastTree 2	Fast approximately-maximum-likelihood phylogenetic trees from nucleotide or protein alignments	Phylogenetic tree of core gene alignments, can be used to assess if the input genomes are closely-related
Prank	Probabilistic multiple alignment program aiming at an evolutionarily correct alignment of closely-related sequences	Alignment of single copy core genes
consambig (EMBOSS)	Creates an ambiguous consensus sequence from a multiple alignment	Build a consensus sequence to identify conserved sequences
GNU parallel	Shell tool for executing jobs in parallel from a input list	Run jobs (Prank, consambig) in parallel
Primer3	Widely used program for the design of PCR primers offering many different input parameters	Design of PCR primers for target sequences
MFEprimer-2.0	Fast primer quality control tool to assess PCR primer specificity	Assess the binding of primer sequences to target and non-target sequences
MPprimer	Design and quality control of multiplex PCR primers	Assess the stability of primer dimers
Mfold	Predicts secondary structures of single stranded nucleic acids	Assess the stability of secondary structures in PCR amplicons

Supplementary table S2

Table S2.2: Primer sequences of the qPCR assays used for the *in vitro* validation, available at https://doi.org/10.7717/peerj.8544/supp-9.

Target species	Gene	Primer	Sequence 5'-3'	Amplicon size (bp)
E. faecalis	*acuI*	Ecfaeca_acuI-F	TCATTTCAAGCATTTACGTTAAGAGA	73
		Ecfaeca_acuI-R	CGTCTAAAGTAATGGTTTCTAGTTGA	
	g3060	Ecfaeca_g3060-F	CCCTCTTTAACATTAATTGGACTTGC	74
		Ecfaeca_g3060-R	TGATTTCTTAATTTATAGGCTTGGCT	
E. faecium	*cysS*	Ecfaeci_cysS-F	GCAGCCACCAATTTACAACGA	86
		Ecfaeci_cysS-R	TCATCTGCCAAATTCTCTGAGG	
	purD	Ecfaeci_purD-F	GGAATGAAAGAACTAGGACGCT	71
		Ecfaeci_purD-R	GGTCCTTCCTTCGTAGCAAT	
P. acidilactici	*asnS*	Pdacidi_asnS-F	GCCATCCACGAATTCCTCCA	80
		Pdacidi_asnS-R	GCACCTTCCGTATCACTAGAAGT	
	g1164	Pdacidi_g1164-F	TTTAGGAGCAATCATCGCAATG	75
		Pdacidi_g1164-R	CCAACTAATACCGCGCTACC	
P. pentosaceus	*nagK*	Pdpento_nagK-F	GGGTGTCACTAACGGAGCAA	107
		Pdpento_nagK-R	CAATTCCGCCCGCTTGTAAG	
	g4364	Pdpento_g4364-F	TGCTTACCACCAACCTGCTT	76
		Pdpento_g4364-R	TGGACGTTCTAGGTTTCCGT	

Chapter 3

Development of a high-throughput microfluidic qPCR system for the quantitative determination of quality-relevant bacteria in cheese

frontiers in Microbiology

ORIGINAL RESEARCH
published: 07 January 2021
doi: 10.3389/fmicb.2020.619166

Development of a High-Throughput Microfluidic qPCR System for the Quantitative Determination of Quality-Relevant Bacteria in Cheese

Matthias Dreier[1,2], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1] and Pilar Junier[2]*

[1] Agroscope, Bern, Switzerland, [2] Laboratory of Microbiology, University of Neuchâtel, Neuchâtel, Switzerland

OPEN ACCESS

Edited by:
Maria Cristina Dantas Vanetti, Universidade Federal de Viçosa, Brazil

Reviewed by:
Mirna Mrkonjic Fuka, University of Zagreb, Croatia
Christophe Monnet, Institut National de la Recherche Agronomique (INRA), France

***Correspondence:**
Matthias Dreier
matthias.dreier@agroscope.admin.ch

Specialty section:
This article was submitted to Food Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 October 2020*
Accepted: *08 December 2020*
Published: *07 January 2021*

Citation:
Dreier M, Berthoud H, Shani N, Wechsler D and Junier P (2021) Development of a High-Throughput Microfluidic qPCR System for the Quantitative Determination of Quality-Relevant Bacteria in Cheese. Front. Microbiol. 11:619166. doi: 10.3389/fmicb.2020.619166

The composition of the cheese microbiome has an important impact on the sensorial quality and safety of cheese. Therefore, much effort has been made to investigate the microbial community composition of cheese. Quantitative real-time polymerase chain reaction (qPCR) is a well-established method for detecting and quantifying bacteria. High-throughput qPCR (HT-qPCR) using microfluidics brings further advantages by providing fast results and by decreasing the cost per sample. We have developed a HT-qPCR approach for the rapid and cost-efficient quantification of microbial species in cheese by designing qPCR assays targeting 24 species/subspecies commonly found in cheese. Primer pairs were evaluated on the Biomark (Fluidigm) microfluidic HT-qPCR system using DNA from single strains and from artificial mock communities. The qPCR assays worked efficiently under identical PCR conditions, and the validation showed satisfying inclusivity, exclusivity, and amplification efficiencies. Preliminary results obtained from the HT-qPCR analysis of DNA samples of model cheeses made with the addition of adjunct cultures confirmed the potential of the microfluidic HT-qPCR system to screen for selected bacterial species in the cheese microbiome. HT-qPCR data of DNA samples of two downgraded commercial cheeses showed that this approach provides valuable information that can help to identify the microbial origin of quality defects. This newly developed HT-qPCR system is a promising approach that will allow simultaneous monitoring of quality-relevant species in fermented foods with high bacterial diversity, thereby opening up new perspectives for the control and assurance of high product quality.

Keywords: real-time qPCR, microbial community composition, microfluidic, cheese quality, cheese microbiome, fermented food, food microbiology, Fluidigm

Abbreviations: BLAST, basic local alignment tool; Cq, quantification cycle; DNA, deoxyribonucleic acid; *E.*, *Enterococcus*; HT-qPCR, high-throughput qPCR; *L.*, *Lactobacillus*; LAB, lactic acid bacteria; NSLAB, non-starter lactic acid bacteria; *Pd.*, *Pediococcus*; *Pr.*, *Propionibacterium*; qPCR, quantitative real-time polymerase chain reaction; *S.*, *Streptococcus*; Tm, melting temperature.

INTRODUCTION

Cheese can be considered a complex ecosystem that is characterized by multiple interactions between its diverse microbial community and environmental conditions. The cheese rind exhibits a high microbial diversity, whereas the composition of the microbiome within the cheese body is less complex (Wolfe et al., 2014; Dugat-Bony et al., 2016). Although the microbiota of raw milk is diverse, several factors, such as pretreatment of the milk, the use of starters, and the thermal conditions applied during cheese making, strongly influence the initial composition of the cheese microbiome. Moreover, the harsh environmental conditions occurring during ripening favor the development of a characteristic ripening microbiota that is especially adapted to an environment characterized by limited levels of fermentable carbohydrates, acidic pH, elevated salt concentrations, and low temperatures (De Filippis et al., 2014; Gobbetti et al., 2018).

The study of the bacterial community composition and of the bacterial population dynamics in cheese has been greatly improved with the advent of culture-independent molecular techniques. Methods such as denaturing gradient gel electrophoresis (DGGE), temporal temperature gradient gel electrophoresis (TTGE), single strand conformational polymorphism (SSCP), length heterogeneity PCR (LH-PCR), and terminal restriction fragment length polymorphism (T-RFLP), were commonly used in the past decades to study the microbial composition of raw milk and cheese, as well as the rind microbiota (Quigley et al., 2011). However, the recent development of next-generation sequencing (NGS) techniques has enabled an even more detailed study of complex microbiomes, and these are now the most widely used approaches in food microbial ecology (De Filippis et al., 2017). The commonest NGS technique used in the analysis of food microbiomes is 16S rRNA gene amplicon-based sequencing, which provides an extensive overview of the food microbiota (Cao et al., 2017). However, identification beyond the genus level is often not possible with this method (Claesson et al., 2010). In the case of cheese and dairy products, species level classification has been achieved by optimization of primer pairs for variable 16S rRNA gene regions, by improved data analysis procedures, and by the establishment of high-quality databases, such as the manually curated DAIRYdb (Meola et al., 2019).

Even though NGS is now routinely used by academic researchers, its use in the food industry is rare. An inherent limitation of the 16S rRNA gene sequencing method is that it only provides the relative abundances of the individual members of the community [operational taxonomic units (OTUs), amplicon sequence variants (ASVs), and taxa]. Complementary approaches, such as quantitative real-time PCR (qPCR) or flow cytometry, are then required to assess the quantitative aspects of the communities (Props et al., 2017). This quantitative analysis is particularly important for fermented foods, as off-flavors may arise due to the abundance of certain microbial populations (Giraffa, 2004). The composition of the cheese microbiome has an important impact on the sensory quality and safety of the final cheese product (Fox et al., 2017). The sensorial quality depends on the microbial biodiversity as well as on the bacterial counts of each individual species (Giraffa, 2004). The metabolic activity of desired and undesired bacterial species is usually sensorially perceivable at counts of $>10^5$ colony-forming units per gram (CFU/g); however, easily noticeable flavor characteristics and off-flavors are typically associated with bacterial counts of 10^6–10^9 CFU/g (Fox et al., 2017).

Quantitative real-time PCR is a well-established method for the detection and quantification of bacteria, such as in pathogen detection in clinical and veterinary diagnostics and in food safety (Curran et al., 2007; Ramirez et al., 2009; Cremonesi et al., 2014; Sartori et al., 2017; Garrido-Maestu et al., 2018). The major limitation of standard qPCR methods is their low throughput, but this has been overcome in recent years with the development of high-throughput qPCR (HT-qPCR) platforms (Ishii et al., 2013; Waseem et al., 2019). HT-qPCR has now been validated and applied to investigate synthetic bacterial soil communities (Kleyer et al., 2017), to determine functional genes in soils (Crane et al., 2018), to quantify pathogens in spiked fecal and environmental water samples (Ishii et al., 2013), to study the gut microbial diversity in piglets (Hermann-Bank et al., 2013), and to quantify dairy *Lactococcus* (*Lc.*) *lactis* and *Leuconostoc* species bacteriophages (Muhammed et al., 2017). However, to our knowledge, HT-qPCR has not yet been used to quantify bacteria in fermented foods, such as cheese. Particularly in the case of raw milk cheeses, microbially induced quality defects, such as off-flavors caused by faulty secondary fermentation or the formation of high quantities of biogenic amines, can frequently lead to a downgrading of cheeses, with significant financial losses. A cost-effective monitoring of desirable and undesirable microorganisms could therefore improve the surveillance of product quality and enable the identification of the causes of microbial cheese defects at an early stage of ripening.

The present study describes the design, validation, and application of a novel microfluidic HT-qPCR system for the simultaneous quantification of multiple bacterial species that are frequently present in raw milk cheeses. We evaluated 24 qPCR assays targeting 23 different bacterial species, including two *Lactococcus lactis* subspecies. The selected target bacteria included lactic acid bacteria (LAB) often used as starters for cheese production, non-starter lactic acid bacteria (NSLAB), and selected species associated with undesired secondary fermentation. A workflow was also developed to facilitate the experimental setup, data filtering, and analysis of the HT-qPCR results. The developed HT-qPCR system was tested under practical conditions by inoculating experimental cheeses with different target species and by including two downgraded commercial cheeses with quality defects in the analysis.

MATERIALS AND METHODS

Selection of Target Species and Primer Design

Twenty-four target species were selected based on a review of the literature and our own preliminary results from 16S

rRNA gene amplicon-based sequencing of Gruyere and Raclette cheeses (unpublished data). The selection criteria were the abundance and frequency of detection, as well as known impacts on cheese quality (**Table 1**). The primer pairs used in this study are listed in **Supplementary Table S1**. New primer pairs were designed for 20 species according to the workflow described in a previous study (Dreier et al., 2020). Briefly, genome assemblies of the target species were downloaded from the National Center for Biotechnology Information (NCBI) and a pan-genome analysis was performed, single copy core genes were selected for primer design and species-specific primer pairs were identified. Three primer pairs were previously published (Dreier et al., 2020). LbhelvF1 and a modified version of LbhelvR1, described elsewhere (Moser et al., 2017), were selected as the primer pair for *Lactobacillus helveticus*. All primers were validated *in silico* by BLAST and Primer-BLAST searches (Johnson et al., 2008; Ye et al., 2012).

Bacterial Strains

For each species, the type strain was selected; for additional strains, isolates from food were preferred. Strains (**Supplementary Data Sheet 1**, target and off-target strain sheets) were obtained from the Agroscope Culture Collection stored at −80°C in sterile reconstituted skim milk powder (10% w/v) and were reactivated and cultivated according to the conditions specified in **Supplementary Data Sheet 1** (cultivation conditions sheet).

DNA Extraction

DNA was extracted from bacterial single strains and from cheese samples, as follows. Bacterial pellets from single strains were harvested from 1 ml overnight cultures by centrifugation (10,000 × g, 5 min, room temperature). 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) SDS was then added to 10 ml of the homogenate, which was then thoroughly mixed and centrifuged (4,000 × g, room temperature, 30 min). The bacterial pellets from the single strains and from the cheese samples were then subjected to a pre-lysis treatment, as described previously (Dreier et al., 2020). Briefly, the pre-lysis treatment included a 15 min incubation in 50 mM sodium hydroxide, followed by an incubation with 2.5 mg/ml lysozyme for 1 h at 37°C. Cell lysis and genomic DNA extraction was performed using the EZ1 DNA Tissue kit and a BioRobot® EZ1 workstation (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Genomic DNA was eluted in a volume of 100 µl and the concentration was measured using a NanoDrop® ND-1000 spectrophotometer

TABLE 1 | Selected species/subspecies and their impact on cheese quality.

Species	Group	Associated defect	Incidence level
Clostridium tyrobutyricum	Raw milk contaminant	Butyric acid fermentation	Species
Enterococcus durans	NSLAB	Biogenic amines (T)	Species
Enterococcus faecalis	NSLAB	Biogenic amines (T)	Species
Enterococcus faecium	NSLAB	Biogenic amines (T)	Species
Levilactobacillus brevis	NSLAB	Biogenic amines (T)	Strain
Lacticaseibacillus casei	NSLAB	–	–
Loigolactobacillus coryniformis	NSLAB	Biogenic amines (H)	Strain
Latilactobacillus curvatus	NSLAB	Biogenic amines (T, P)	Strain
Lactobacillus delbrueckii	Starter	–	–
Limosilactobacillus fermentum	NSLAB/(Whey starter)	(Excess gas formation)	Species
Lactobacillus helveticus	Starter/Adjunct	–	–
Lentilactobacillus parabuchneri	NSLAB	Biogenic amines (H)	Strain
Lacticaseibacillus paracasei	NSLAB/Adjunct	–	–
Lactiplantibacillus paraplantarum	NSLAB	–	–
Lactiplantibacillus plantarum	NSLAB	–	–
Lacticaseibacillus rhamnosus	NSLAB/Adjunct	–	–
Latilactobacillus sakei	NSLAB	–	–
Lactococcus lactis subsp. cremoris	Starter	–	–
Lactococcus lactis subsp. lactis	Starter	–	–
Leuconostoc mesenteroides	Starter/Adjunct	–	–
Pediococcus acidilactici	NSLAB	–	–
Pediococcus pentosaceus	NSLAB	–	–
Propionibacterium freudenreichii	Adjunct/Raw milk	Propionic acid fermentation	Species
Streptococcus thermophilus	Starter	–	–

T, Tyramine; P, Putrescine; H, Histamine.

(NanoDrop Technologies, Thermo Fisher Scientific, Waltham, MA, United States).

Reagents and Conditions for Standard qPCR

The inclusivity of the primer pairs was assessed by performing qPCR with 2 ng DNA of 2–34 strains of the target species in technical duplicates (**Supplementary Data Sheet 1**, target strains). The qPCR assays were performed in a total reaction mix volume of 12 µl, containing 6 µl 2× SsoFast™ EvaGreen® Supermix with low ROX (Biorad, Cressier, Switzerland), 500 nM of forward and reverse primers, and 2 µl of DNA. The qPCR cycling conditions consisted in an initial denaturation at 95°C for 1 min, followed by 35 cycles of 95°C for 5 s and 60°C for 1 min. The melting curve analysis was performed using a gradient from 60 to 95°C, with 1°C steps per 3 s. All qPCR assays were run on a Corbett Rotor-Gene 3000 (Qiagen). Rotor-Gene 6000 Software 1.7 was used for analysis, with dynamic tube normalization and a threshold of 0.05 for quantification cycle (Cq) value calculation; the five first cycles were ignored for the determination of the Cq values. The peak calling threshold for the melt curve analysis was set to −2 dF/dT, and the temperature threshold was set at 2°C lower than the positive control peak.

Preamplification of DNA Samples

An assay mix was prepared by pooling 1 µl of each primer (100 µM) in a total volume of 200 µl DNA suspension buffer [10 mM tris(hydroxymethyl)aminomethane, 0.1 mM ethylenediaminetetraacetic acid, pH 8]. A volume of 1.25 µl DNA sample was mixed with 3.75 µl preamplification pre-mix consisting of 2.5 µl 2× TaqMan PreAmp Master Mix (Thermo Fisher Scientific, Waltham, MA, United States), 0.5 µl of pooled assay mix, and 0.75 µl DNase-free water. Preamplification was performed using a Labcycler (SensoQuest, Göttingen, Germany) thermal cycler using the following conditions: an initial denaturation step at 95°C for 10 min, followed by 14 cycles at 95°C for 15 s and 60°C for 4 min. The preamplification primers were eliminated from the reactions by treating the samples with 2 µl diluted Exonuclease I (4 U/µl, Thermo Fisher Scientific, Waltham, MA, United States) at 37°C for 30 min, followed by enzyme inactivation at 80°C for 15 min. The final reactions were diluted 10-fold with DNA suspension buffer and stored at -20°C.

Microfluidic HT-qPCR

HT-qPCR was performed using a 192.24 Dynamic Array integrated fluidic circuit (IFC; Fluidigm Corporation, San Francisco, CA, United States). DNA samples from pure bacterial cultures were diluted to 3 ng/µl prior to qPCR measurement. 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 (Fluidigm, 2015). 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 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), consisting of 24 double stranded target species sequences separated by thymine spacers five base pairs in length. A map representation of the HT-qPCR standard is shown in **Figure 1**, and the sequence is available in **Supplementary Data Sheet 2**. The dried gBlock gene fragment pellet (Molecular weight: 1440635.7 u) was resuspended with DNA suspension buffer [10 mM tris(hydroxymethyl)aminomethane, 0.1 mM ethylenediaminetetraacetic acid, pH 8] at a concentration of 10 ng/µl. Copy numbers were calculated using the following equation:

$$10\frac{ng}{l} \times 0.69\frac{fmol}{ng} \times 1 \times 10^{-15}\frac{mol}{fmol} \times 6.022 \times 10^{23}\frac{copies}{mol}$$

$$= 4.16 \times 10^{9}\ copies/l$$

HT-qPCR Standard Calibration Curves

Copy numbers for quantification were calculated using duplicate standard calibration curves ranging from 10^8 to 10^3 copies/µl (**Supplementary Figure S1** in **Supplementary Data Sheet 3**).

HT-qPCR Samples

We assessed the specificity of the primer pairs using DNA from pure cultures of 84 strains (**Supplementary Data Sheet 1**, off-target strains sheet). For each strain, the cultivation from stock culture and the DNA extraction were performed twice independently. With the exception of *Leuconostoc mesenteroides* (four strains) and *Lacticaseibacillus casei* (two strains), three strains of each target species were selected. In addition, we also selected DNA of 12 type strains of species often occurring in dairy products or closely related to one of the target species (*Leuconostoc carnosum* and *Streptococcus salivarius*). The HT-qPCR was performed with DNA samples diluted to 3 ng/µl.

A mock community consisting of the type strains of the 24 target species/subspecies at concentrations of about 1×10^6 copies/µl was also prepared (**Supplementary Data Sheet 1**, Mock community sheet). The DNA concentration for the corresponding number of genome copies was estimated by taking the genome size of the type strain, if available. Otherwise, we used the average genome size[1] and an average weight of 1.096×10^{-21} g per base pair. A 10-fold dilution series

[1]https://www.ncbi.nlm.nih.gov/genome

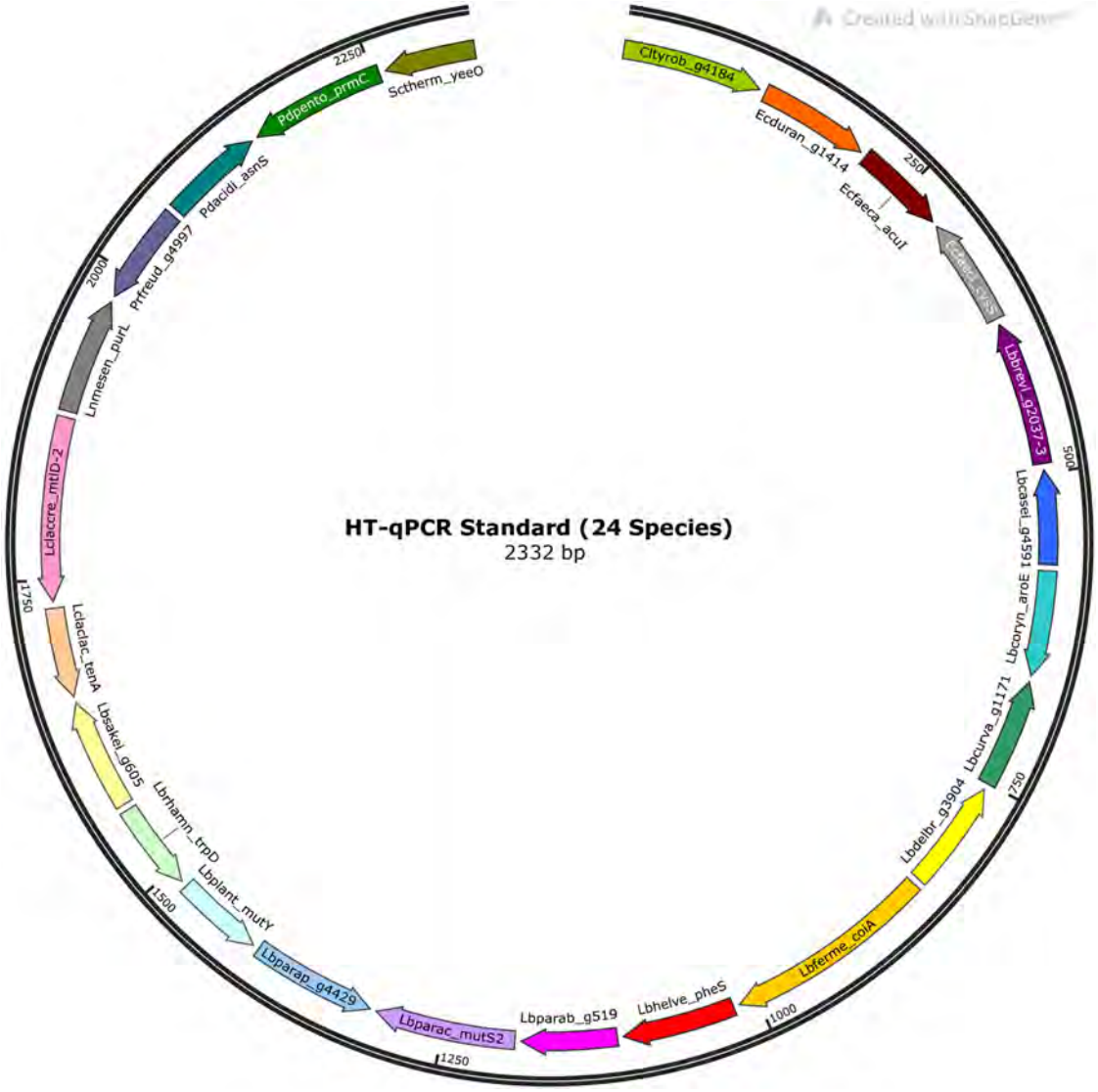


FIGURE 1 | Map representation of the HT-qPCR standard. The standard is a linear, double-stranded DNA molecule 2,332 base pairs in length and consisting of target sequences of 24 species separated by thymine spacers 5 base pairs in length. The map was created using SnapGene Viewer 5.0.7.

of the mock community was prepared and subjected to preamplification to enrich the target sequences in the mock community dilutions (10^4–10 copies/μl). Mock community dilutions without preamplification (10^5–10^2 copies/μl) were also measured.

192.24 Dynamic Array IFC Setup

The validation was performed on multiple 192.24 Dynamic Array IFCs. All samples (pure bacterial culture DNAs, no template controls, mock community, and HT-qPCR standard dilution series) were included, and eight primer pairs were measured in triplicate in each run.

Production of Model Cheeses With Adjunct Cultures

Fifteen model cheeses with adjunct cultures of selected target species [*Levilactobacillus brevis, L. casei, Loigolactobacillus coryniformis, Latilactobacillus curvatus, Limosilactobacillus fermentum, L. helveticus, Lentilactobacillus parabuchneri, Lacticaseibacillus paracasei, Lactiplantibacillus plantarum,*

Lacticaseibacillus rhamnosus, *Latilactobacillus sakei, Leuconostoc mesenteroides, Pediococcus* (*Pd.*) *acidilactici, Pd. pentosaceus*, and *Propionibacterium* (*Pr.*) *freudenreichii*] and 4 control cheeses (without adjunct cultures) were produced in the experimental cheese dairy at Agroscope (Bern, Switzerland) on four different days. The experimental design for the production of the 19 model cheeses and the conditions used for the preparation of the 15 adjunct cultures are listed in **Supplementary Table S2**. The pasteurized vat milk was inoculated by centrifuging 50 ml of each adjunct culture (4,000 × g, room temperature, 10 min) and resuspending in 50 ml sterile reconstituted skim milk powder (10% w/v) before addition to the milk. The estimated concentration of adjunct culture in the milk vat was 10^4–10^5 CFU/ml.

The Raclette-type semi-hard model cheeses were produced from 50 l of pasteurized milk, using a combination of the mesophilic starter RSW 901 (*Lc. lactis* ssp. *lactis*, *Lc. lactis* ssp. *cremoris*, *Lc. lactis* ssp. *diacetylactis*) and the mixed mesophilic/thermophilic starter MK 401 (*Lc. lactis* ssp. *lactis*, *Lactobacillus delbrueckii* subsp. *lactis* and *S. thermophilus*) (Liebefeld Kulturen AG, Switzerland). The milk was pre-ripened at 28–32°C for 30 min, followed by rennet addition and coagulation for 25 min at 32°C, and then cutting and stirring at 32°C for 25 min. The temperature was then increased to 36°C for 10 min and the milk was stirred for a further 35 min. The whey-curd mixture was filled into molds and pressed for 4 h at 34°C, 4 h at 32°C, and finally 8 h at 28°C. The cheeses (30 cm in diameter, about 6 kg) were immersed in a 20% (w/w) saline solution (11–13°C, 14 h), and smear-ripened in a maturing cellar (10–11°C, 90–96% relative humidity) for up to 120 days. The samples were collected after 111, 113, 118, and 120 days of ripening for the cheeses manufactured on days 1, 2, 3, and 4, respectively.

HT-qPCR Application on Cheese Samples

We calculated the copy numbers for quantification using standard calibration curves ranging from 10^7 to 10^3 copies/µl (**Supplementary Figure S2** in **Supplementary Data Sheet 3**). The measurement was performed on a single 192.24 Dynamic Array IFC. All samples (HT-qPCR standard dilutions, no template controls, and cheese samples) were measured in technical triplicates.

Data Analysis

Results from the 192.24 Dynamic Array IFCs for the validation runs were combined for the analysis with the Fluidigm Real-Time PCR Analysis Software version 4.5.2 (Fluidigm Corp.). The quality threshold was set to 0.5, the quantification cycle (Cq) threshold for all reactions was set to 0.05, and the baseline correction was set to constant. The settings used for the melting curve analysis were: a peak sensitivity of 3 and a peak ratio threshold of 0.8, the qPCR assay-specific peak detection ranges are available in **Supplementary Table S3**. The melting curve peak threshold was set to 0.05 −dRn/dT for the validation runs and to 0.025 −dRn/dT for the run with the cheese samples, based on visual inspection of the baseline fluorescence. The Real-Time PCR Analysis Software flags all reactions that do not conform to the selected thresholds (i.e., low quality score, multiple or no melting curve peaks, or reactions where the normalized fluorescence is below the threshold). The data were then exported to a csv file. A python script (biomarkdataparser.py) was used to filter the data and to calculate the number of copies/µl in the samples based on the calibration curves. 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 (Fluidigm, 2018). Average copies/µl were only calculated if at least two of three reactions were positive; otherwise, the results were interpreted as negative. The raw data (csv export) from the Real-Time PCR Analysis Software, the biomarkdataparser.py script and the jupyter-notebooks used to make the figures are available in **Supplementary Data Sheet 4** and on GitHub[2].

Analysis of Volatile Carboxylic Acids and Biogenic Amines

Volatile carboxylic acids in cheese were esterified with ethanol, and analyzed by gas chromatography as described by Fröhlich-Wyder et al. (2013) using a Hewlett Packard HP 6890 gas chromatograph (Agilent Technologies, Basel, Switzerland) equipped with a Hewlett Packard Ultra 2 cross linked phenyl methyl silicone fused silica capillary column (50 m, 0.32 mm, 0.52 mm) and a flame ionization detector (FID). Biogenic amines in cheese were analyzed as described by Ascone et al. (2017) using a UPLC system (UltiMate 3000 RS; Thermo Fisher Scientific, Reinach, Switzerland) equipped with a C18 column (Accucore C18: 2.6 mm, 150 × 4.6 mm; Thermo Fisher Scientific, Reinach, Switzerland). All measurements were carried out in duplicate.

RESULTS

Specificity of the qPCR Assays

The inclusivity of the qPCR assays was assessed by performing standard qPCR with DNA from single strains of each target species (**Supplementary Data Sheet 5**). The inclusivity was 100% for all the tested qPCR assays (**Table 2**). The qPCR assay for *L. casei* was only tested with two *L. casei* strains, due to the limited availability of these strains in public strain collections. The *in silico* validation of the primer pair showed that all available genomes of *L. casei* and *L. zeae* contain a perfectly matching target sequence, in contrast to genomes of any other species of the *Lactobacillaceae* family (NCBI:txid33958) available in the NCBI Microbial Genomes BLAST database, including complete and draft genomes (as of July 2020; data not shown).

The specificity of the qPCR assays was assessed by performing HT-qPCR with DNA from single strains of two to four strains of the target and selected type strains of off-target species. The raw Cq data showed high quantification cycles for several off-target reactions, mainly for the qPCR assays for the detection

[2]https://github.com/biologger/htqpcr_validation_data

TABLE 2 | Standard qPCR results of inclusivity assessment.

Species	Mean Cq	SD	Mean T_m	SD	Inclusivity
Clostridium tyrobutyricum	14.32	1.05	80.02	0.14	25/25
Enterococcus durans	14.5	1.27	81.3	0.22	25/25
Enterococcus faecalis	14.55	0.71	75.5	0.0	22/22
Enterococcus faecium	11.95	0.53	80.8	0.11	25/25
Levilactobacillus brevis	14.65	2.21	82.55	0.11	18/18
Lacticaseibacillus casei	14.36	0.41	83.4	0.12	2/2
Loigolactobacillus coryniformis	12.95	0.4	83.19	0.16	19/19
Latilactobacillus curvatus	13.4	0.58	82.21	0.09	25/25
Lactobacillus delbrueckii	14.62	1.15	83.52	0.09	34/34
Limosilactobacillus fermentum	14.19	1.29	87.12	0.1	24/24
Lactobacillus helveticus	14.42	1.32	78.42	0.13	24/24
Lentilactobacillus parabuchneri	13.98	1.45	81.56	0.29	25/25
Lacticaseibacillus paracasei	13.88	1.06	84.34	0.23	21/21
Lactiplantibacillus paraplantarum	14.04	1.02	84.21	0.23	14/14
Lactiplantibacillus plantarum	13.86	0.67	76.94	0.1	24/24
Lacticaseibacillus rhamnosus	13.93	1.42	81.92	0.48	24/24
Latilactobacillus sakei	11.98	0.45	83.0	0.06	24/24
Lactococcus lactis subsp. cremoris	14.28	1.18	81.23	0.17	25/25
Lactococcus lactis subsp. lactis	14.53	0.87	80.02	0.2	25/25
Leuconostoc mesenteroides	13.68	1.14	82.28	0.32	23/23
Pediococcus acidilactici	11.61	0.46	78.89	0.22	20/20
Pediococcus pentosaceus	12.64	0.96	78.48	0.1	25/25
Propionibacterium freudenreichii	15.1	0.85	85.5	0.03	25/25
Streptococcus thermophilus	13.73	0.82	78.89	0.24	25/25

Mean values and standard deviation (SD) of quantification cycle (Cq) and melting temperature (T_m) of inclusivity assessment by standard qPCR for strains of the target species. Raw data is available in ***Supplementary Data Sheet 5****.*

of *L. delbrueckii*, *Lc. lactis subsp. lactis,* and *S. thermophilus* (**Supplementary Figure S3** in **Supplementary Data Sheet 3**). Background noise was reduced by applying two filter criteria to the data: all reactions flagged by the analysis software and all reactions with fewer than 800 copies/µl were interpreted as negative reactions, as recommended by the manufacturer (**Supplementary Figure S4** in **Supplementary Data Sheet 3**). All target species strains were detected by HT-qPCR, and only one cross-reaction was detected with the filtered average Cq values (**Figure 2**). The cross-reaction of the *Lactiplantibacillus paraplantarum* assay with the off-target strain *L. coryniformis* (DSM 20004) was only detected in one of two different DNA extracts, and the Cq value was about eight cycles higher than that for the *L. paraplantarum* DNA samples. In summary, the *L. paraplantarum* assay had a specificity of 0.9939, while all other tested assays were specific.

Sensitivity and Dynamic Range of the qPCR Assays

The qPCR assay performance was assessed with a 10-fold dilution series of the qPCR standard consisting of all 24 target sequences in a range from 10^8 to 10^3 copies/µl. The calculated efficiency of the qPCR assays ranged between 87 and 97%. The linear regression equations ($Cq\ slope * log$[copies] intercept) had slopes between −3.39 and −3.68 and correlation coefficients between 0.992 and 0.998. The sensitivity of the assays without preamplification is given by the cut-off Cq value corresponding to 800 copies/µl (Cq 23.8–26.4), as calculated using the linear equations of the standard calibration curves (**Supplementary Figure S1** in **Supplementary Data Sheet 3**).

We validated the quantification of the targets in mixtures by HT-qPCR analysis of samples of a 10-fold dilution series of a mock community consisting of DNA from 24 type strains in a range between 10^5 and 10^2 copies/µl. All targets were detected in the diluted mock community sample containing 10^4 copies/µl, and 14 of 24 assays detected the target a dilution of 10^3 copies/µl (**Figure 3**). The concentrations of target DNA in the mock community were calculated based on the initial DNA concentration of the single strain sample and the genome size of the target species (**Supplementary Data Sheet 1**, Mock community sheet). The predicted concentrations were compared to the measured copies/µl (**Supplementary Figure S5** in **Supplementary Data Sheet 3**). The assays for the detection of *L. brevis, L. sakei*, *L. paracasei*, *Pr. freudenreichii,* and *S. thermophilus* had lower initial concentrations of the target sequence than predicted. By contrast, the assays for *Enterococcus* (*E.*) *durans*, *E. faecalis*, *Lc. lactis subsp. lactis*, *Pd. acidilactici,* and *Pd. pentosaceus* showed similar values for the predicted and measured number of copies/µl, though the assays did not detect the target in the diluted sample containing 10^3 copies/µl, whereas the 14 other assays did.

Preamplification Efficiency

The increase in sensitivity due to preamplification reactions was assessed by preamplification of a 10-fold dilution series of a mock community consisting of DNA from 24 type strains in a range between 10^4 and 10 copies/µl and subsequent HT-qPCR analysis. All species were detected down to a dilution of 10^2 copies/µl in the pre-amplified mock community sample, whereas in the samples with the highest dilution of 10 copies/µl, 14 of 24 targets were detected (**Figure 3**). The efficiency of the preamplification reaction for the qPCR assays was assessed by comparing the Cq values obtained for a diluted mock community sample (10^4 copies/µl) with preamplification to Cq values without preamplification (**Table 3**). The Cq values for the sample with preamplification decreased, on average, by 7.43 cycles (range

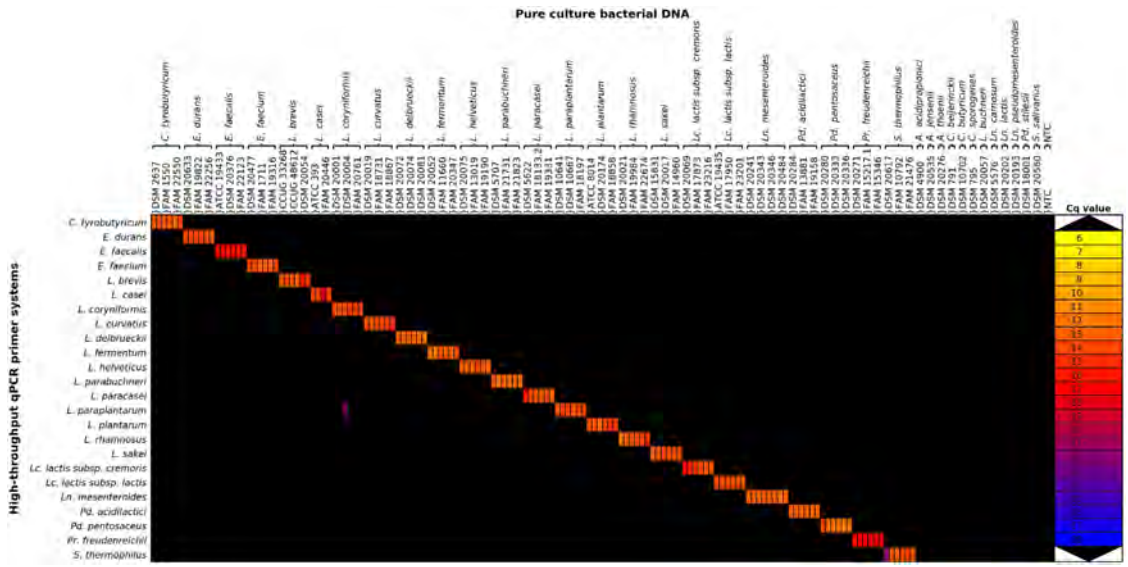


FIGURE 2 | Heatmap of filtered average quantification cycle data of the exclusivity assessment. Average Cq values of the filtered Cq data are represented, but only data points where at least two of the three technical replicates were positive are depicted. Raw Cq data and filtered Cq data heatmaps are available in **Supplementary Figures S3, S4** in **Supplementary Data Sheet 3**. *A.*, *Acidipropionibacterium*; *C.*, *Clostridium*; *E.*, *Enterococcus*; *L.*, *Lactobacillus*; *Lc.*, *Lactococcus*; *Ln.*, *Leuconostoc*; *Pd.*, *Pediococcus*; *Pr.*, *Propionibacterium*; *S.*, *Streptococcus*; NTC, no template control.

6.49–9.85) compared to the Cq values for the sample without preamplification.

Application of HT-qPCR to Raclette-Type Model Cheese

The ability of the HT-qPCR system to quantify the target species in real cheese DNA samples was verified by manufacturing 19 Raclette-type model cheeses with the target species adjuncts. The DNA extracts from 19 Raclette-type model cheeses were analyzed by HT-qPCR (**Figure 4**). The volatile carboxylic acids and biogenic amines of the cheeses were also analyzed, as these metabolites are often elevated in defective cheeses and serve as indicators of the presence of undesirable microorganisms.

The four starter LAB species (*Lc. lactis* subsp. *lactis*, *Lc. lactis* subsp. *cremoris*, *L. delbrueckii,* and *S. thermophilus*) were detected in all cheese DNA samples. All 15 adjunct culture species were detected in the corresponding cheese DNA sample, except for sample S07, where no *L. helveticus* was detected. Low concentrations of *L. helveticus* were detected in sample S06, indicating that the adjunct culture with *L. helveticus* had mistakenly been added to the wrong cheese vat. In several samples, cross-contaminations of target species from cheeses that were produced on the same production day were detected at distinct lower concentrations. The concentration of propionic acid was elevated in cheese samples S13 (14.93 mmol/kg) and S18 (37.83 mmol/kg) and, to a lesser extent, in cheese S17 (4.53 mmol/kg) and S19 (2.5 mmol/kg). Increased amounts of tyramine were measured in cheese samples S2 (171.78 mg/kg), S3 (284.67 mg/kg), S5 (482.33 mg/kg), and S19 (155.72 mg/kg), while in samples S5 and S8, the concentration of putrescine (292.8 mg/kg) and histamine (320.35 mg/kg) were increased, respectively.

Application of HT-qPCR to Downgraded Commercial Cheeses With Quality Defects

The potential of the HT-qPCR system to identify the microbial causes of cheese defects was demonstrated by HT-qPCR analysis of DNA extracts from two commercial cheese samples with quality defects (the C1 alpine cheese and C2 Raclette cheese, **Figure 5**).

The alpine cheese sample (C1) had increased concentrations of propionic acid (36.5 mmol/kg) and biogenic amines, mainly histamine (733 mg/kg) and to a lesser extent tyramine (398 mg/kg) and putrescine (417 mg/kg). The cheese contained high concentrations of the typical thermophilic starter species *S. thermophilus,* whereas *L. delbrueckii. L. helveticus,* and *L. parabuchneri* were present at concentrations over 10^5 copies/µl and *L. coryniformis*, *L. curvatus*, *L. paracasei,* and *Pd. pentosaceus* had concentrations between 10^4 and 10^5 copies/µl. Low concentrations ($<10^4$ copies/µl) of *E. faecalis* and *L. paraplantarum* were detected.

The Raclette cheese sample (C2) had elevated levels of biogenic amines, mainly tyramine (545 mg/kg), but also histamine (185 mg/kg). Both subspecies of *Lc. lactis* and *Leuconostoc mesenteroides* used in mesophilic starters were detected, with *Lc. lactis subsp. lactis* as the predominant species at more than 10^6 copies/µl. *L. helveticus* and *L. parabuchneri*

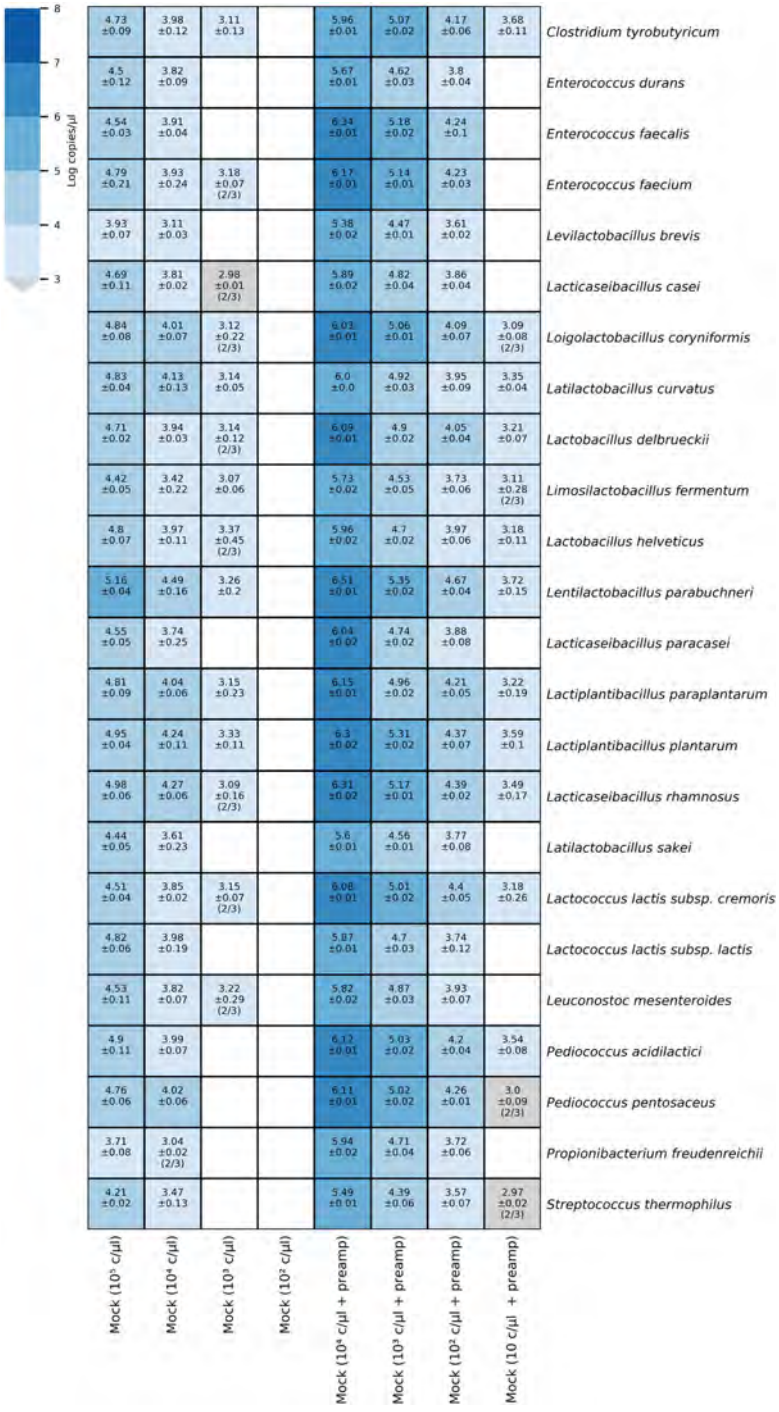


FIGURE 3 | Heatmap of microfluidic qPCR results for mock community dilutions. The heatmap annotation depicts the average logarithmic copies/µl and the standard deviation. When not all samples were positive, the number of positive samples out of the total number of samples (triplicates) is given in brackets. On the left side, the results correspond to the diluted mock community DNA samples quantified without preamplification whereas, the diluted mock community DNA samples depicted on the right side were pre-amplified.

TABLE 3 | Quantification cycle values of a mock community sample with and without preamplification.

Species	Cq before preamplification	Cq after preamplification	ΔCq
Clostridium tyrobutyricum	20.71	13.67	7.05
Enterococcus durans	21.51	15.02	6.49
Enterococcus faecalis	22.89	14.43	8.45
Enterococcus faecium	21.10	13.32	7.78
Levilactobacillus brevis	23.42	15.51	7.90
Lacticaseibacillus casei	21.70	14.40	7.30
Loigolactobacillus coryniformis	20.65	13.56	7.09
Latilactobacillus curvatus	20.68	14.20	6.49
Lactobacillus delbrueckii	20.53	13.13	7.40
Limosilactobacillus fermentum	22.51	14.28	8.23
Lactobacillus helveticus	21.07	14.08	6.99
Lentilactobacillus parabuchneri	19.35	12.22	7.12
Lacticaseibacillus paracasei	21.21	13.24	7.97
Lactiplantibacillus paraplantarum	20.22	12.86	7.36
Lactiplantibacillus plantarum	20.96	13.60	7.36
Lacticaseibacillus rhamnosus	20.15	12.96	7.19
Latilactobacillus sakei	21.86	14.99	6.87
Lactococcus lactis subsp. cremoris	22.42	14.20	8.22
Lactococcus lactis subsp. lactis	20.85	14.32	6.53
Leuconostoc mesenteroides	21.00	13.96	7.04
Pediococcus acidilactici	20.55	13.15	7.40
Pediococcus pentosaceus	20.05	12.76	7.29
Propionibacterium freudenreichii	23.31	13.46	9.85
Streptococcus thermophilus	22.04	15.20	6.84

were found at concentrations of about 10^5 copies/µl, whereas *L. paracasei*, *L. rhamnosus,* and *Pr. freudenreichii* were present at concentrations below 10^4 copies/µl.

DISCUSSION

Validation of qPCR Assays and the Microfluidic HT-qPCR System

The qPCR assays validated in this study were highly specific. However, the qPCR assay for *L. casei* is not able to differentiate between *L. casei* and *L. zeae* species, two similar species for which a reclassification was recently proposed (Huang et al., 2020). The false positive cross-reaction of the *L. paraplantarum* assay in one *L. coryniformis* DNA sample was most likely due to a cross-contamination. A similar melting curve peak and the negative result of an independent DNA extraction from the same pure cultured strain support this assumption.

Background fluorescence signals in the raw data were mainly caused by three qPCR assays, specifically the assays for *L. delbrueckii*, *Lc. lactis* subsp. *lactis*, and *S. thermophilus*. Background signals may occur due to weak amplification of primer dimers in samples without the target sequence. The qPCR assay-specific cut-off values (equivalent to 800 copies/µl) were calculated from standard calibration curves and used to reduce background signals. Measures to increase the signal to the background ratio have also been reported in other microfluidic HT-qPCR studies. For example, a previous study (Ishii et al., 2013) reported that some (TaqMan) probes had to be redesigned because the probes failed to obtain sufficiently strong signals to separate them from background signals. Another study (Hermann-Bank et al., 2013) developed the Gut Microbiotassay on a 48 × 48 Access Array (Fluidigm Corp.) and excluded Cq values exceeding primer-specific cut-off values during data analysis.

The sensitivity of the tested qPCR assays was limited by the nanoliter-scale reactions used in the microfluidic qPCR system. This limitation for microfluidic HT-qPCR can be addressed by adding a preamplification step as a part of the experimental workflow. Preamplification increased the sensitivity of all assays in the mock communities. However, the delta Cq values calculated from target sequences and pre-amplified target sequences differed considerably for the 24 qPCR assays; consequently, the Cq data from samples with pre-amplification do not allow a reliable quantitative analysis and can therefore, only be used for qualitative detection of targets.

Application of Microfluidic HT-qPCR to Cheese Samples

Given the technical limit of 800 copies/µl for qPCR reactions, the theoretical limit of detection of the assays was calculated as 8×10^4 genome equivalents/g cheese. However, it should be noted that for culture-independent quantitative methods, the DNA extraction method can have a significant impact on the results obtained. It is known that residues from the food matrix such as fats, proteins and calcium in DNA samples can inhibit subsequent PCR reactions (Wilson, 1997). DNA extraction can also have an influence on the recovery rates of different bacteria, e.g., due to the different composition of cell walls and the resulting differences in the efficiency of cell lysis, such as between Gram-positive and Gram-negative bacteria (Quigley et al., 2012). Starter LAB grow very fast during cheese production, typically reaching counts of $>10^8$ CFU/g within the first 24 h. By contrast, the growth of NSLAB is significantly slower and occurs mainly during the first weeks of ripening, reaching bacterial counts of 10^6–10^8 CFU/g, depending on the species (Fox et al., 2017). Quantitative studies have shown that the population density of species relevant for the organoleptic quality of cheese typically ranges from 10^6 to 10^{10} genome equivalents/g cheese (Falentin et al., 2010, 2012; Turgay et al., 2011; Desfosses-Foucault et al., 2012; Moser et al., 2018). At lower population densities, the formation of metabolites is too low to be reliably perceived by

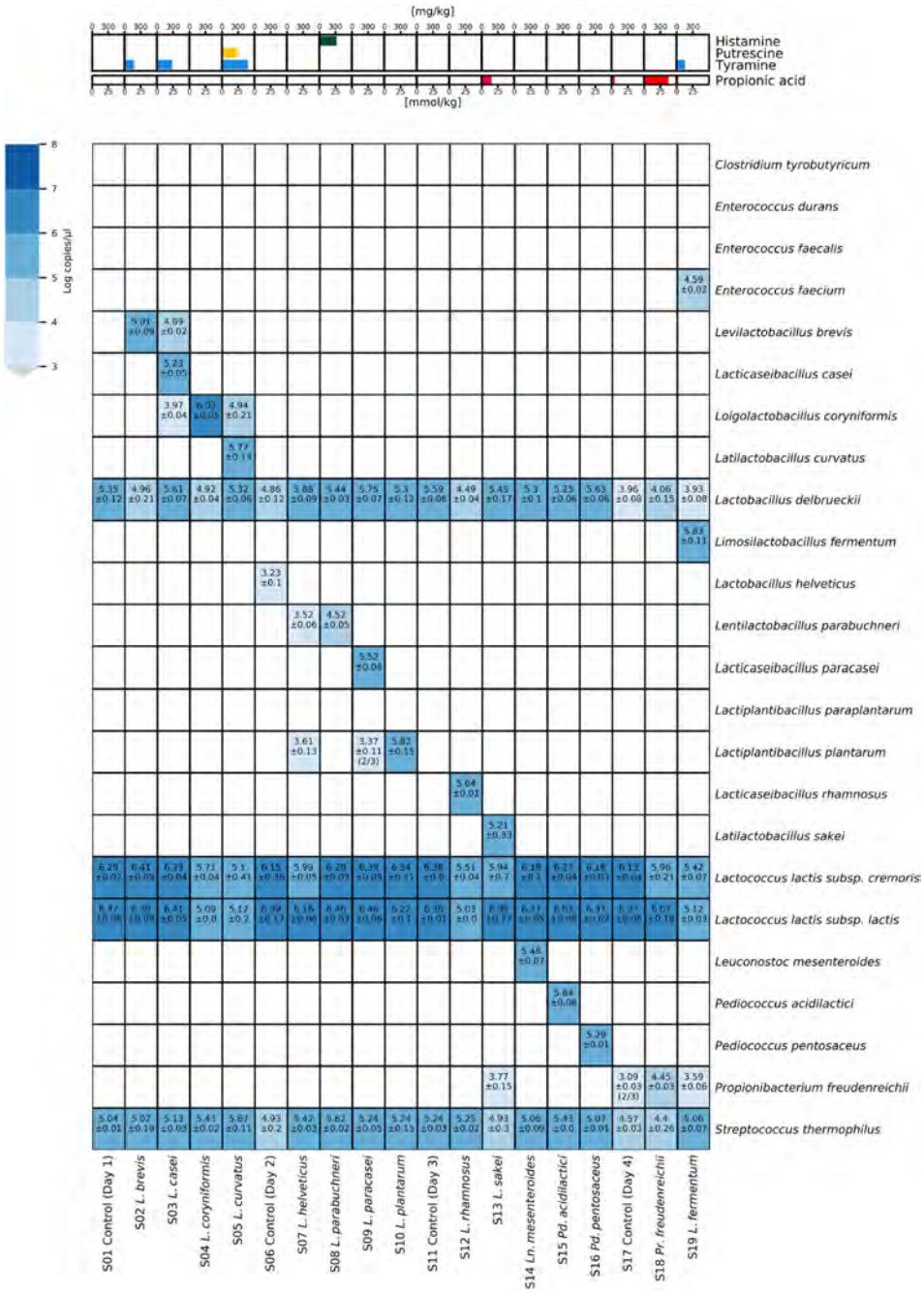


FIGURE 4 | Heatmap of microfluidic HT-qPCR results of Raclette-type model cheese samples produced with adjunct cultures. The results of the chemical analysis of the model cheese are shown in the upper panel, while the results of the HT-qPCR analysis (without preamplification) are shown in a heatmap in the lower panel.

sensory perception. The results obtained from the HT-qPCR analysis of the mock community dilutions indicate that all assays are able to quantify a minimal population density of 10^6 genome equivalents/g cheese. Despite this rather high detection limit, HT-qPCR would still be a valuable tool for cost-effective monitoring of species relevant for the sensory quality of cheese. In addition,

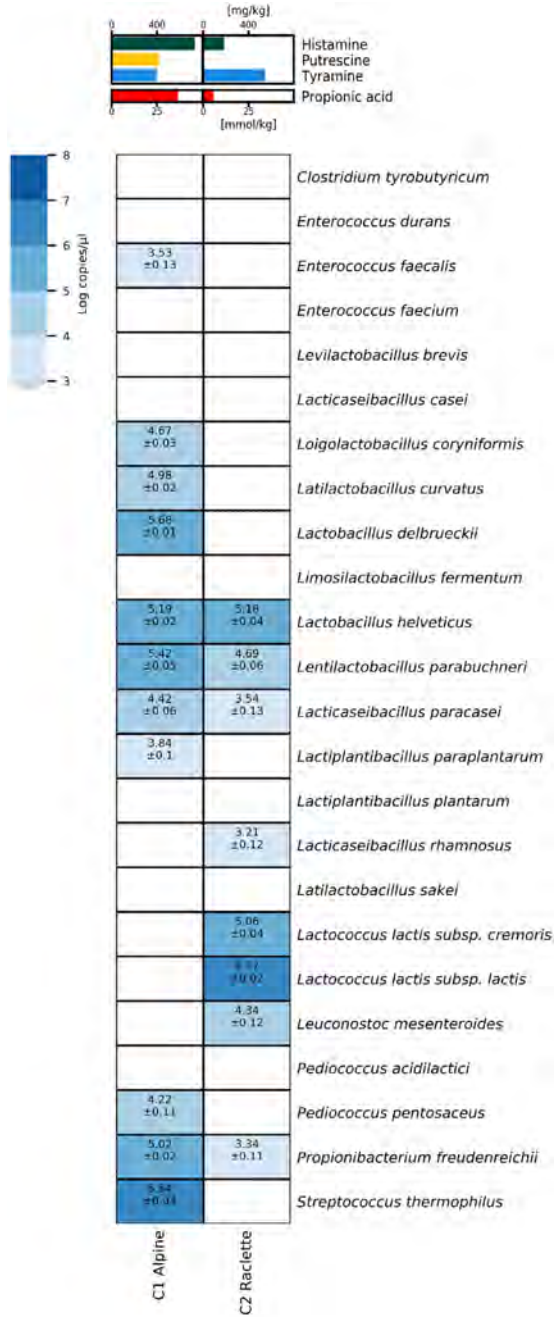


FIGURE 5 | Heatmap of microfluidic qPCR results (without preamplification) for commercial cheese samples with defects.

detection of species with lower abundancies (e.g., NSLAB species) in early stages of ripening could optionally be achieved using a preamplification step.

The application of the HT-qPCR system to model and commercial cheese samples was used to show the potential of the new method to detect a broad range of quality-relevant species in cheese samples, including starter LAB, NSLAB, and raw milk-associated contaminants that may cause severe cheese defects during ripening. The application of the microfluidic qPCR assays on model cheeses with adjunct cultures of selected target species confirmed the successful detection and quantification of these target species in cheese DNA samples. In addition, we observed the presence of bacteria that had not been deliberately added with the adjunct cultures in several cheese samples. These cross-contaminations most likely originated from equipment used in parallel during the simultaneous production of the experimental cheeses on the same day (e.g., cheese harps used for cutting the curd and the system used for filling the curd/whey mixture into the cheese molds). However, the unexpected presence of *Pr. freudenreichii* in sample S13 remains unexplained, as no adjunct culture with *Pr. freudenreichii* was used on that production day. The growth of *Pr. freudenreichii* in the cheese in sample S13 resulted in a similarly increased concentration of propionic acid (14.9 mmol/kg) as in other cheeses (S17, S18, and S19) in which *Pr. freudenreichii* was detected (**Figure 4**). Studies examining the environment and production facilities of cheese dairies show that bacteria present in raw milk and cheese are quite abundant and can persist on surfaces, despite frequent cleaning (Somers et al., 2001; Bokulich and Mills, 2013; Stellato et al., 2015). The source of the *E. faecium* contamination in S19 was identified as a contaminated stock culture of one of the used *L. fermentum* strains, as confirmed by partial 16S rRNA gene sequencing of single-colony DNA (**Supplementary Data Sheet 6**).

The selection of qPCR assays designed for the HT-qPCR system included species of undesirable bacteria found in raw milk. Various microbiologically induced quality defects in cheese are related to contamination of the processed milk with undesirable bacteria. The most common microbial causes of cheese defects are faulty fermentations, such as butyric acid fermentation (typically caused by *Clostridium tyrobutyricum*) and propionic acid fermentation (typically caused by *Pr. freudenreichii*), and the formation of biogenic amines (Bachmann et al., 2011). Tyramine, histamine, cadaverine, putrescine, and β-phenylethylamine (PEA) are the most abundant biogenic amines in cheese (Linares et al., 2011). Various NSLAB species play an important role in the excessive formation of biogenic amines in cheese (Barbieri et al., 2019). The formation of biogenic amines is a strain-specific characteristic of various NSLAB species. For example, strains of *L. parabuchneri* have been repeatedly isolated from cheeses heavily contaminated with histamine, whereas aminogenic strains of *E. faecium* are often present in cheeses with elevated tyramine content. Similarly, strains of *L. curvatus* have been shown to be potent producers of tyramine and putrescine (Benkerroum, 2016; Diaz et al., 2016; Wüthrich et al., 2017). The determination of metabolites like volatile carboxylic acids and biogenic amines often provides helpful information that clarifies the microbial origin of faulty fermentations and other cheese defects. However, the simultaneous quantitative determination of undesirable bacterial species using HT-qPCR opens up new perspectives for an efficient and cost-effective

diagnosis of the causes of microbially induced cheese defects. Notably, the early and reliable detection of the microbial causes of cheese defects is an important precondition for tracing the sources of contamination and taking corrective actions.

In the model cheese experiments, samples with elevated tyramine content contained either *L. brevis*, *L. curvatus,* or *E. faecium*; all three species are known tyramine producers (Coton and Coton, 2009; Bunkova et al., 2010; Ladero et al., 2012). Sample S05 containing *L. curvatus* also showed elevated levels of putrescine, while sample S08 containing *L. parabuchneri* had elevated levels of histamine.

In the alpine cheese (sample C1), the histamine concentration was strongly increased (733 mg/kg), and an increased population density (5.42 log copies/µl) of *L. parabuchneri* was detected. The additional presence of *E. faecalis* and *L. curvatus* likely explains the formation of tyramine and putrescine. Moreover, the increased concentration of propionic acid correlates with the increased numbers of *Pr. freudenreichii* detected in this sample.

Similarly, the detection of *L. parabuchneri* most likely explains the increased concentration of histamine in the defective commercial Raclette cheese (sample C2). However, the results of the HT-qPCR analysis did not allow identification of a species that could account for the elevated tyramine content. In all likelihood, a species not covered by our qPCR assays was responsible for the high concentrations of tyramine. Strains of several *Lactobacillus* species other than the NSLAB species targeted here have been reported to produce tyramine (Bunkova et al., 2010; Benkerroum, 2016).

The setup of the method described here allows the exchange or extension of the qPCR assays for the detection of additional species or functional genes (e.g., the *hdc* gene, important in histamine production). Furthermore, the outlined workflow allows an efficient validation of new primer pairs for integration into the HT-qPCR system. We demonstrated here the potential of the HT-qPCR system to quantify simultaneously multiple bacterial species in cheese DNA samples. However, this approach could also be of interest for the investigation of other fermented foods such as kimchi, sauerkraut or sausages that also contain complex microbial compositions which include to some extent the same LAB species as present in cheese (Plengvidhya et al., 2007; Cocolin et al., 2009; Jung et al., 2011).

The HT-qPCR approach presented in this study offers a fast and affordable simultaneous quantitative screening of 24 species/subspecies relevant for the quality of cheese. A single 192.24 Dynamic Array IFC chip enables the screening of 56 cheese DNA samples in technical triplicates from 24 species/subspecies in several hours. Moreover, the developed script for data cleaning and visualization then allows immediate visualization and interpretation of the data exported from the Fluidigm Real-Time PCR Analysis software, thereby facilitating the rapid interpretation of the data. The high sample capacity of the microfluidic high-throughput system and the high specificity of the qPCR assays are key factors required for fast, accurate, and cost-efficient monitoring of desired and undesired microorganisms affecting sensory cheese quality.

Another advantage is that the system can easily be expanded with additional assays to cover further product-specific species or to adapt the system to other fermented products. Preliminary results from model cheeses and downgraded commercial cheeses showed that the application of HT-qPCR to complex fermented products such as cheese could be of interest for identification of the microbiological causes of sensorially perceivable quality defects. Particularly in the production of raw milk cheese, the application of HT-qPCR could be very useful for monitoring the composition of the ripening microbiota, thereby ensuring a constant product quality.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MD, HB, NS, DW, and PJ conceived and designed the experiments, authored or reviewed drafts of the manuscript, and approved the final draft. MD performed the experiments, analyzed the data, and prepared figures and tables. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Daniel Marzohl, Nadine Sidler, Elvira Wagner, and Kotchanoot Srikham for their valuable technical help. The HT-qPCR data produced and analyzed in this manuscript were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zürich. We especially thank Aria Minder and Silvia Kobel for their technical support in conducting the analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.619166/full#supplementary-material

REFERENCES

Ascone, P., Maurer, J., Haldemann, J., Irmler, S., Berthoud, H., Portmann, R., et al. (2017). Prevalence and diversity of histamine-forming *Lactobacillus parabuchneri* strains in raw milk and cheese – a case study. *Int. Dairy J.* 70, 26–33. doi: 10.1016/j.idairyj.2016.11.012

Bachmann, H. P., Fröhlich-Wyder, M. T., Jakob, E., Roth, E., Wechsler, D., Beuvier, E., et al. (2011). "Cheese/Raw milk cheeses," in *Encyclopedia of Dairy Sciences*, 2nd Edn, ed. J. W. Fuquay (San Diego, CA: Academic Press), 652–660. doi: 10.1016/b978-0-12-374407-4.00519-7

Barbieri, F., Montanari, C., Gardini, F., and Tabanelli, G. (2019). Biogenic amine production by lactic acid bacteria: a review. *Foods* 8:17. doi: 10.3390/foods8010017

Benkerroum, N. (2016). Biogenic amines in dairy products: origin, incidence, and control means. *Compr. Rev. Food Sci. Food Saf.* 15, 801–826. doi: 10.1111/1541-4337.12212

Bokulich, N. A., and Mills, D. A. (2013). Facility-specific "house" microbiome drives microbial landscapes of artisan cheesemaking plants. *Appl. Environ. Microbiol.* 79, 5214–5223. doi: 10.1128/AEM.00934-13

Bunkova, L., Bunka, F., Mantlova, G., Cablova, A., Sedlacek, I., Svec, P., et al. (2010). The effect of ripening and storage conditions on the distribution of tyramine, putrescine and cadaverine in Edam-cheese. *Food Microbiol.* 27, 880–888. doi: 10.1016/j.fm.2010.04.014

Cao, Y., Fanning, S., Proos, S., Jordan, K., and Srikumar, S. (2017). A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Front. Microbiol.* 8:1829. doi: 10.3389/fmicb.2017.01829

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38:e200. doi: 10.1093/nar/gkq873

Cocolin, L., Dolci, P., Rantsiou, K., Urso, R., Cantoni, C., and Comi, G. (2009). Lactic acid bacteria ecology of three traditional fermented sausages produced in the North of Italy as determined by molecular methods. *Meat Sci.* 82, 125–132. doi: 10.1016/j.meatsci.2009.01.004

Coton, E., and Coton, M. (2009). Evidence of horizontal transfer as origin of strain to strain variation of the tyramine production trait in *Lactobacillus brevis*. *Food Microbiol.* 26, 52–57. doi: 10.1016/j.fm.2008.07.009

Crane, S. L., van Dorst, J., Hose, G. C., King, C. K., and Ferrari, B. C. (2018). Microfluidic qPCR enables high throughput quantification of microbial functional genes but requires strict curation of primers. *Front. Environ. Sci.* 6:145. doi: 10.3389/fenvs.2018.00145

Cremonesi, P., Pisani, L. F., Lecchi, C., Ceciliani, F., Martino, P., Bonastre, A. S., et al. (2014). Development of 23 individual TaqMan(R) real-time PCR assays for identifying common foodborne pathogens using a single set of amplification conditions. *Food Microbiol.* 43, 35–40. doi: 10.1016/j.fm.2014.04.007

Curran, T., Coyle, P. V., McManus, T. E., Kidney, J., and Coulter, W. A. (2007). Evaluation of real-time PCR for the detection and quantification of bacteria in chronic obstructive pulmonary disease. *FEMS Immunol. Med. Microbiol.* 50, 112–118. doi: 10.1111/j.1574-695X.2007.00241.x

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

De Filippis, F., Parente, E., and Ercolini, D. (2017). Metagenomics insights into food fermentations. *Microb. Biotechnol.* 10, 91–102. doi: 10.1111/1751-7915.12421

Desfosses-Foucault, E., Dussault-Lepage, V., Le Boucher, C., Savard, P., Lapointe, G., and Roy, D. (2012). Assessment of probiotic viability during cheddar cheese manufacture and ripening using propidium monoazide-PCR quantification. *Front. Microbiol.* 3:350. doi: 10.3389/fmicb.2012.00350

Diaz, M., Ladero, V., Redruello, B., Sanchez-Llana, E., del Rio, B., Fernandez, M., et al. (2016). A PCR-DGGE method for the identification of histamine-producing bacteria in cheese. *Food Control* 63, 216–223. doi: 10.1016/j.foodcont.2015.11.035

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

Dugat-Bony, E., Garnier, L., Denonfoux, J., Ferreira, S., Sarthou, A. S., Bonnarme, P., et al. (2016). Highlighting the microbial diversity of 12 French cheese varieties. *Int. J. Food Microbiol.* 238, 265–273. doi: 10.1016/j.ijfoodmicro.2016.09.026

Falentin, H., Henaff, N., Le Bivic, P., Deutsch, S. M., Parayre, S., Richoux, R., et al. (2012). Reverse transcription quantitative PCR revealed persistency of thermophilic lactic acid bacteria metabolic activity until the end of the ripening of Emmental cheese. *Food Microbiol.* 29, 132–140. doi: 10.1016/j.fm.2011.09.009

Falentin, H., Postollec, F., Parayre, S., Henaff, N., Le Bivic, P., Richoux, R., et al. (2010). Specific metabolic activity of ripening bacteria quantified by real-time reverse transcription PCR throughout Emmental cheese manufacture. *Int. J. Food Microbiol.* 144, 10–19. doi: 10.1016/j.ijfoodmicro.2010.06.003

Fluidigm (2015). *Fluidigm Gene Expression with the 192.24 IFC Using Delta Gene Assay Quick Reference*. Available Online at: 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 (accessed July 23, 2020).

Fluidigm (2018). *Real-Time PCR Analysis User Guide*. Available Online at: https://www.fluidigm.com/binaries/content/documents/fluidigm/search/hippo%3Aresultset/real-time-pcr-analysis-ug-68000088/fluidigm%3Afile (accessed July 31, 2020).

Fox, P. F., Guinee, T. P., Cogan, T. M., and McSweeney, P. L. H. (2017). "Microbiology of cheese ripening," in *Fundamentals of Cheese Science*, eds P.F. Fox, T. P. Guinee, T. M. Cogan, P. L. H. McSweeney (Boston, MA: Springer), 333–390. doi: 10.1007/978-1-4899-7681-9_11

Fröhlich-Wyder, M.-T., Guggisberg, D., Badertscher, R., Wechsler, D., Wittwer, A., and Irmler, S. (2013). The effect of *Lactobacillus buchneri* and *Lactobacillus parabuchneri* on the eye formation of semi-hard cheese. *Int. Dairy J.* 33, 120–128. doi: 10.1016/j.idairyj.2013.03.004

Garrido-Maestu, A., Azinheiro, S., Carvalho, J., and Prado, M. (2018). Rapid and sensitive detection of viable *Listeria monocytogenes* in food products by a filtration-based protocol and qPCR. *Food Microbiol.* 73, 254–263. doi: 10.1016/j.fm.2018.02.004

Giraffa, G. (2004). Studying the dynamics of microbial populations during food fermentation. *FEMS Microbiol. Rev.* 28, 251–260. doi: 10.1016/j.femsre.2003.10.005

Gobbetti, M., Di Cagno, R., Calasso, M., Neviani, E., Fox, P. F., and De Angelis, M. (2018). Drivers that establish and assembly the lactic acid bacteria biota in cheeses. *Trends Food Sci. Technol.* 78, 244–254. doi: 10.1016/j.tifs.2018.06.010

Hermann-Bank, M. L., Skovgaard, K., Stockmarr, A., Larsen, N., and Mølbak, L. (2013). The Gut microbiotassay: a high-throughput qPCR approach combinable with next generation sequencing to study gut microbial diversity. *BMC Genomics* 14:788. doi: 10.1186/1471-2164-14-788

Huang, C. H., Chen, C. C., Liou, J. S., Lee, A. Y., Blom, J., Lin, Y. C., et al. (2020). Genome-based reclassification of *Lactobacillus casei*: emended classification and description of the species *Lactobacillus zeae*. *Int. J. Syst. Evol. Microbiol.* 70, 3755–3762. doi: 10.1099/ijsem.0.003969

Ishii, S., Segawa, T., and Okabe, S. (2013). Simultaneous quantification of multiple food- and waterborne pathogens by use of microfluidic quantitative PCR. *Appl. Environ. Microbiol.* 79, 2891–2898. doi: 10.1128/AEM.00205-13

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201

Jung, J. Y., Lee, S. H., Kim, J. M., Park, M. S., Bae, J. W., Hahn, Y., et al. (2011). Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl. Environ. Microbiol.* 77, 2264–2274. doi: 10.1128/AEM.02157-10

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

Ladero, V., Fernandez, M., Calles-Enriquez, M., Sanchez-Llana, E., Canedo, E., Martin, M. C., et al. (2012). Is the production of the biogenic amines tyramine and putrescine a species-level trait in enterococci? *Food Microbiol.* 30, 132–138. doi: 10.1016/j.fm.2011.12.016

Linares, D. M., Martin, M. C., Ladero, V., Alvarez, M. A., and Fernandez, M. (2011). Biogenic amines in dairy products. *Crit. Rev. Food Sci. Nutr.* 51, 691–703. doi: 10.1080/10408398.2011.582813

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

Moser, A., Berthoud, H., Eugster, E., Meile, L., and Irmler, S. (2017). Detection and enumeration of *Lactobacillus helveticus* in dairy products. *Int. Dairy J.* 68, 52–59. doi: 10.1016/j.idairyj.2016.12.007

Moser, A., Schafroth, K., Meile, L., Egger, L., Badertscher, R., and Irmler, S. (2018). Population dynamics of *Lactobacillus helveticus* in Swiss Gruyère-type cheese manufactured with natural whey cultures. *Front. Microbiol.* 9:637. doi: 10.3389/fmicb.2018.00637

Muhammed, M. K., Krych, L., Nielsen, D. S., and Vogensen, F. K. (2017). A high-throughput qPCR system for simultaneous quantitative detection of dairy

Lactococcus lactis and *Leuconostoc* bacteriophages. *PLoS One* 12:e0174223. doi: 10.1371/journal.pone.0174223

Plengvidhya, V., Breidt, F. Jr., Lu, Z., and Fleming, H. P. (2007). DNA fingerprinting of lactic acid bacteria in sauerkraut fermentations. *Appl. Environ. Microbiol.* 73, 7697–7702. doi: 10.1128/AEM.01342-07

Props, R., Kerckhof, F. M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., et al. (2017). Absolute quantification of microbial taxon abundances. *ISME J.* 11, 584–587. doi: 10.1038/ismej.2016.117

Quigley, L., O'Sullivan, O., Beresford, T. P., Paul Ross, R., Fitzgerald, G. F., and Cotter, P. D. (2012). A comparison of methods used to extract bacterial DNA from raw milk and raw milk cheese. *J. Appl. Microbiol.* 113, 96–105. doi: 10.1111/j.1365-2672.2012.05294.x

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., and Cotter, P. D. (2011). Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese. *Int. J. Food Microbiol.* 150, 81–94. doi: 10.1016/j.ijfoodmicro.2011.08.001

Ramirez, M., Castro, C., Palomares, J. C., Torres, M. J., Aller, A. I., Ruiz, M., et al. (2009). Molecular detection and identification of *Aspergillus* spp. from clinical samples using real-time PCR. *Mycoses* 52, 129–134. doi: 10.1111/j.1439-0507.2008.01548.x

Sartori, C., Boss, R., Ivanovic, I., and Graber, H. U. (2017). Development of a new real-time quantitative PCR assay for the detection of *Staphylococcus aureus* genotype B in cow milk, targeting the new gene *adlb*. *J. Dairy Sci.* 100, 7834–7845. doi: 10.3168/jds.2017-12820

Somers, E. B., Johnson, M. E., and Wong, A. C. L. (2001). Biofilm formation and contamination of cheese by nonstarter lactic acid bacteria in the dairy environment. *J. Dairy Sci.* 84, 1926–1936. doi: 10.3168/jds.S0022-0302(01)74634-6

Stellato, G., De Filippis, F., La Storia, A., and Ercolini, D. (2015). Coexistence of lactic acid bacteria and potential spoilage microbiota in a dairy processing environment. *Appl. Environ. Microbiol.* 81, 7893–7904. doi: 10.1128/AEM.02294-15

Turgay, M., Irmler, S., Isolini, D., Amrein, R., Fröhlich-Wyder, M.-T., Berthoud, H., et al. (2011). Biodiversity, dynamics, and characteristics of *Propionibacterium freudenreichii* in Swiss Emmentaler PDO cheese. *Dairy Sci. Technol.* 91, 471–489. doi: 10.1007/s13594-011-0024-7

Waseem, H., Jameel, S., Ali, J., Saleem Ur Rehman, H., Tauseef, I., Farooq, U., et al. (2019). Contributions and challenges of high throughput qPCR for determining antimicrobial resistance in the environment: a critical review. *Molecules* 24:163. doi: 10.3390/molecules24010163

Wilson, I. G. (1997). Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* 63, 3741–3751. doi: 10.1128/AEM.63.10.3741-3751.1997

Wolfe, B. E., Button, J. E., Santarelli, M., and Dutton, R. J. (2014). Cheese rind communities provide tractable systems for *in situ* and *in vitro* studies of microbial diversity. *Cell* 158, 422–433. doi: 10.1016/j.cell.2014.05.041

Wüthrich, D., Berthoud, H., Wechsler, D., Eugster, E., Irmler, S., and Bruggmann, R. (2017). The histidine decarboxylase gene cluster of *Lactobacillus parabuchneri* was gained by horizontal gene transfer and is mobile within the species. *Front. Microbiol.* 8:218. doi: 10.3389/fmicb.2017.00218

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134. doi: 10.1186/1471-2105-13-134

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.7 Supplementary Information

The supplementary material for this arcticle is available at https://www.frontiersin.org/articles/10.3389/fmicb.2020.619166/full#supplementary-material.

Supplementary tables

Table S3.1: Primer pairs used in the study (continued on next page)

Species	Primer	Sequence 5'-3'	Amplicon size (bp)	Reference
Clostridium tyrobutyricum	Cltyrob_g4184-F	TGTCCATCTTATCATCAATGGGT	111	This study
	Cltyrob_g4184-R	CACCATTGCCTAAATCTCCAAGA		
Enterococcus durans	Ecduran_g1414-F	ACACTTGGATCACATTAGGATTGA	88	This study
	Ecduran_g1414-R	TCATAGCTTAACTTCCTCGTAACT		
Enterococcus faecalis	Ecfaeca_acuI-F	TCATTTCAAGCATTTACGTTAAGAGA	73	(Dreier et al. 2020)
	Ecfaeca_acuI-R	CGTCTAAAGTAATGGTTTCTAGTTGA		
Enterococcus faecium	Ecfaeci_cysS-F	GCAGCCACCAATTTACAACGA	86	(Dreier et al. 2020)
	Ecfaeci_cysS-R	TCATCTGCCAAATTCTCTGAGG		
Levilactobacillus brevis	Lbbrevi_g2037-3F	TGAGGCCAGTAGTACGGTT	114	This study
	Lbbrevi_g2037-3R	CTACCAGAGTCCATCGCATG		
Lacticaseibacillus casei	Lbcasei_g4591-F	GCCTCACCAAGTATCAGACACT	74	This study
	Lbcasei_g4591-R	CACATTGTGTTCATGGAGCTG		
Loigolactobacillus coryniformis	Lbcoryn_aroE-F	GACAGCTCGGACATCCTACG	83	This study
	Lbcoryn_aroE-R	AGCGCCAGTCCATAGTTTGT		
Latilactobacillus curvatus	Lbcurva_g1171-F	CAACTTTGATGCGTGAGTTAGA	86	This study
	Lbcurva_g1171-R	TCTATTTCGTTCGTATTCAGCCT		
Lactobacillus delbrueckii	Lbdelbr_g3904-F	AAGAAGAAGCGCCCTACCTG	87	This study
	Lbdelbr_g3904-R	GCTGCCAGGCTCTAGTAAAGT		
Limosilactobacillus fermentum	Lbferme_coiA-F	GGTTATCAGCAAGTAGGAGTGG	167	This study
	Lbferme_coiA-R	ATTGGCTGTAGGTAGTCTAACTG		
Lactobacillus helveticus	LbhelvF1 (pheS)	AGGTTCAAAGCATCCAATCAATATT	89	(Moser et al. 2017) / This study
	LbhelvR (pheS)	TTTCGGGACCTTGCACTACTTTA		
Lentilactobacillus parabuchneri	Lbparab_g519-F	TGGATCTGAAATGTCGCATCG	75	This study
	Lbparab_g519-R	AGTGTTGACCAGGTGTAATGT		
Lacticaseibacillus paracasei	Lbparac_mutS2-F	GGCAGGTTCAACTAGGCATTT	111	This study
	Lbparac_mutS2-R	CGCCACTTACTTTGACAACCG		
Lactiplantibacillus paraplantarum	Lbparap_g4429-F	TTTATCGCGCTGCACTGAAC	99	This study
	Lbparap_g4429-R	TTAGGCCGAGCTTCAGAACG		

Lactiplantibacillus plantarum	Lbplant_mutY-F Lbplant_mutY-R	TGATGATTACCAACCTGAAGAACT CTGGAGCTGATATTCACGTTTAAA	70	This study
Lacticaseibacillus rhamnosus	Lbrhamn_trpD-F Lbrhamn_trpD-R	CTGGGAAGGCACGTGAAGAATTA CATGCCACCACATCCTTACGTTT	70	This study
Latilactobacillus sakei	Lbsakei_g605-F Lbsakei_g605-R	TCAGTCGATTAAGAAGGTAGTGGT ATTGTGAATGACCGTTTGTGG	88	This study
Lactococcus lactis subsp. cremoris	Lclaccre_mtlD-2F Lclaccre_mtlD-2R	GGAAGGTTTACAGCACAAAGAGATT AGCAAAGCGAGAGATAATCATTTGA	145	This study
Lactococcus lactis subsp. lactis	Lclaclac_tenA-F Lclaclac_tenA-R	GGTCGCCCAATCCTTTGTATC TGCGATTTGGACGGTAGCTT	71	This study
Leuconostoc mesenteroides	Lnmesen_purL-F Lnmesen_purL-R	ATGATGAGGACGAGTTAGCAG TCCACCTTCAGATAAATCATGCG	92	This study
Pediococcus acidilactici	Pdacidi_asnS-F Pdacidi_asnS-R	GCCATCCACGAATTCCTCCA GCACCTTCCGTATCACTAGAAGT	80	(Dreier et al. 2020)
Pediococcus pentosaceus	Pdpento_prmC-F Pdpento_prmC-R	TGGTTATCGTCAAGGTACGGA ACCCAAAGAATTCGATCGTGT	108	This study
Propionibacterium freudenreichii	Prfreud_g4997-F Prfreud_g4997-R	ATCCTCCCGCCCTACTTCTT GGCCAGTCCCATGTGAATGA	78	This study
Streptococcus thermophilus	Sctherm_yeeO-F Sctherm_yeeO-R	TCGTCTTGGGTCTACTCTCCA TCACAGATTGCTCAGTTCCCA	72	This study

References

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

Moser, A., Berthoud, H., Eugster, E., Meile, L., and Irmler, S. (2017). Detection and enumeration of Lactobacillus helveticus in dairy products. International Dairy Journal 68, 52-59. doi: 10.1016/j.idairyj.2016.12.007.

Table S3.2: Adjunct cultures for model cheeses

DNA Sample No.	Adjunct culture	Strains	Medium	Cultivation conditions (Temperature, Duration, Condition)	Production day	Cheese No.
S01	Control (day 1)				1	1
S02	*Levilactobacillus brevis*	DSM 20054^T FAM 23126	MRS broth	37°C, 1 d, aerobic	1	3
S03	*Lacticaseibacillus casei*	ATCC 393^T FAM 20446	MRS broth	30°C, 1 d, aerobic	1	5
S04	*Loigolactobacillus coryniformis*	DSM 20001^T FAM 20761	MRS broth	30°C, 1 d, aerobic	1	6
S05	*Latilactobacillus curvatus*	DSM 20019^T FAM 18731	MRS broth	30°C, 1 d, aerobic	1	8
S06	Control (day 2)				2	9
S07	*Lactobacillus helveticus*	DSM 20075 FAM 19190	MRS broth	37°C, 1 d, aerobic	2	11
S08	*Lentilactobacillus parabuchneri*	DSM 5707^T FAM 21731	MRS broth	37°C, 2 d, aerobic	2	12
S09	*Lacticaseibacillus paracasei*	DSM 5622 FAM 18133	MRS broth	30°C, 1 d, aerobic	2	14
S10	*Lactiplantibacillus plantarum*	DSM 20174^T FAM 20642	MRS broth	30°C, 1 d, aerobic	2	16
S11	Control (day 3)				3	17
S12	*Lacticaseibacillus rhamnosus*	DSM 20021 FAM 19984	MRS broth	30°C, 1 d, aerobic	3	18
S13	*Latilactobacillus sakei*	DSM 20017^T FAM 14960	MRS broth	30°C, 1 d, aerobic	3	19
S14	*Leuconostoc mesenteroides*	DSM 20346 FAM 24147	MRS broth	30°C, 1 d, aerobic	3	21
S15	*Pediococcus acidilactici*	DSM 20284^T FAM 19158	MRS broth	37°C, 1 d, aerobic	3	22
S16	*Pediococcus pentosaceus*	DSM 20336^T FAM 19161	MRS broth	37°C, 1 d, aerobic	3	23
S17	Control (day 4)				4	25
S18	*Propionibacterium freudenreichii*	DSM 20271^T FAM 15217.1	MRS-lactose broth	30°C, 3 d, anaerobic	4	28
S19	*Limosilactobacillus fermentum*	DSM 20052^T FAM 20347	MRS broth	37°C, 1 d, aerobic	4	32

MRS: De Man, J.C., M. Rogosa, and M.E. Sharpe. A medium for the cultivation of lactobacilli. Journal of Applied Bacteriology, 1960. 23(1): p. 130-135.
MRS-lactose: MRS with 2 % lactose instead of 2% glucose

Table S3.3: Peak detection ranges for the melting curve analysis by the Real-Time PCR Analysis Software (Fluidigm)

Primer pair	Min	Max
Cltyrob_g4184	76.50	80.00
Ecduran_g1414	77.00	81.00
Ecfaeca_acuI	72.00	75.00
Ecfaeci_cysS	77.00	81.00
Lbbrevi_g2037-3	79.00	83.00
Lbcasei_g4591	80.00	84.00
Lbcoryn_aroE	79.00	83.00
Lbcurva_g1171	79.00	83.00
Lbdelbr_g3904	81.00	85.00
Lbferme_coiA	84.00	88.00
Lbhelve_pheS	75.00	79.00
Lbparab_g519	78.00	82.00
Lbparac_mutS2	80.50	84.50
Lbparap_g4429	80.00	84.00
Lbplant_mutY	73.00	77.00
Lbrhamn_trpD	79.00	83.00
Lbsakei_g605	79.00	83.00
Lclaccre_mtlD-2	78.00	82.00
Lclaclac_tenA	77.00	81.00
Lnmesen_purL	79.00	83.00
Pdacidi_asnS	76.00	79.00
Pdpento_prmC	75.00	79.00
Prfreud_g4997	83.00	86.00
Sctherm_yeeO	75.00	79.00

Supplementary data sheet 1

Bacterial strains used in the study

The data sheet 1 is a large Excel workbook containing multiple worksheets with tabular data. The data set is too large to include it in the thesis. A short description of the content is given below.

Sheets:

1. List of target strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.
2. List of off-target strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.
3. Cultivation conditions for the strains from the Agroscope Culture Collection used for the *in vitro* validation of primer specificity.
4. Estimated genome size of type strains and composition of the mock community.

Supplementary data sheet 2

Fasta sequence of the HT-qPCR standard

> HT-qPCR_standard (24 target sequences)
TGTCCATCTTATCATCAATGGGTGGAATAGGTGGGCTCGGAGGCGGATCAAGAAATGATAACATTG
GTTCAATTTTAAATAACTTAAATCTTGGAGATTTAGGCAATGGTGTTTTTACACTTGGATCACATTA
GGATTGAGTATTGCTTTGATCCTCCTTGCTTGGTTGCTTGTACGAGGAGTTACGAGGAAGTTAAGCT
ATGATTTTTTCATTTCAAGCATTTACGTTAAGAGAACAAGATACATCTGTTATTGGTCAACTAGAAA
CCATTACTTTAGACGTTTTTTCATCTGCCAAATTCTCTGAGGCAGTTTCTTGGCGGAATTTCAAATTG
TCAAACGTATTCTTCAGTCGTTGTAAATTGGTGGCTGCTTTTTCTACCAGAGTCCATCGCATGGCTCT
TTAATTTTATGAACATTGGATCCTTAACGTTTTTTAGCGGCTGATACCACTCATGCCTAAGTGACTTA
GGAACCGTACTACTGGCCTCATTTTTCACATTGTGTTCATGGAGCTGGCTGTGTTGGGCATGGGGGC
CAGTGACTTCAAGTGTCTGATACTTGGTGAGGCTTTTTGACAGCTCGGACATCCTACGTATAATGGA
TTGCCGATGCTGATCCAGCAAGGTGTCTTAGCTTACAAACTATGGACTGGCGCTTTTTTTCTATTTCG
TTCGTATTCAGCCTGTCCTTGCCCAACTTGCATGCCAATTTGATTGACCCGTTCTTCTAACTCACGCA
TCAAAGTTGTTTTTGCTGCCAGGCTCTAGTAAAGTAGTAGGCCAGCATGGTCAAGTCGTAGTTGCTG
GAATTGTAGCCCATCAGGTAGGGCGCTTCTTCTTTTTTTGGTTATCAGCAAGTAGGAGTGGCGGTTC
GTTGGCTACTGGGTCCAACTTACCAAAAGCGGCTCCTCCACCGTGAAATCCAAGCGCGTTTTACCCA
ACTAGTTAATAACCGGCCCAGTTTAACCTTCTGGGAGCCCCACCGAGCACAGTTAGACTACCTACA
GCCAATTTTTTAGGTTCAAAGCATCCAATCAATATTATTTTGGATGATTTGGAAAGCTACTTCATCG
GTATGGGTTATAAAGTAGTGCAAGGTCCCGAAATTTTTTGGATCTGAAATGTCGCATCGAACAGGC
CGATAACTTAATCGAAACCTCGACCTACATTACACCTGGTCAACACTTTTTTGGCAGGTTCAACTAG
GCATTTTAAAAATGAAGGTACCAACTGATGAGCTGGAAAAAATCAAGCCAAGTAAGCAATCAGCG
GCGCAACGGCCGGTTGTCAAAGTAAGTGGCGTTTTTTTAGGCCGAGCTTCAGAACGTAAATTTTGG
CAATGACATAACCAGAACAAGTTATCAAGACCAGTGCGTTGTCGTGACGGTTCAGTGCAGCGCGAT
AAATTTTTCTGGAGCTGATATTCACGTTTAAAATAATTTTCTACCGCCGTAACTAGTTCTTCAGGTTG
GTAATCATCATTTTTCATGCCACCACATCCTTACGTTTGGCAGAAAAGGCGAGCAGGTGATTTAATT
CTTCACGTGCCTTCCCAGTTTTTTCAGTCGATTAAGAAGGTAGTGGTAAAATCCGATGCACTAACAA
CTTTAGCGCAAGCGGGTGTGACCCCACAAACGGTCATTCACAATTTTTTTGCGATTTGGACGGTAGC
TTCTTCACTCGCATAGGTTTCAATCCATTTTTGATACAAAGGATTGGGCGACCTTTTTAGCAAAGCG
AGAGATAATCATTTGATGATAGCTCATCAAATCTTCTTCCGTAAAATATTGGGGCCATTTGGCTAAA
AGTAAAGCGCGTGTTTCTTTCTGTACACCCTTTAAAGCACTTAAAATCTCTTTGTGCTGTAAACCTTC
CTTTTTATGATGAGGACGAGTTAGCAGCGCAACAGGTCATTCTAGATGTCACTGATCAACGTCTGCT
TAATAGTGCGCATGATTTATCTGAAGGTGGATTTTTGGCCAGTCCCATGTGAATGATGACTGCGGCA
CCGAGCACGATGCGGAAGAGCGTAAGGAAGAAGTAGGGCGGGAGGATTTTTTGCCATCCACGAAT
TCCTCCAAAAGAACAACTTTGTCTACGTTAACACTCCGTTGATTACTTCTAGTGATACGGAAGGTGC
TTTTTACCCAAAGAATTCGATCGTGTTGACTAAGATCTTTTATAACTTCCGCTTGGTCAATATACTTC
TTTTCTAAGAAATACTTTTTTAGGTCCGTACCTTGACGATAACCATTTTTTCACAGATTGCTCAGTTC
CCAGAAGATTAAGAATGGCTCCACCAAAAAAAATGGAGAGTAGACCCAAGACGA

Supplementary data sheet 3

Supplementary Figures

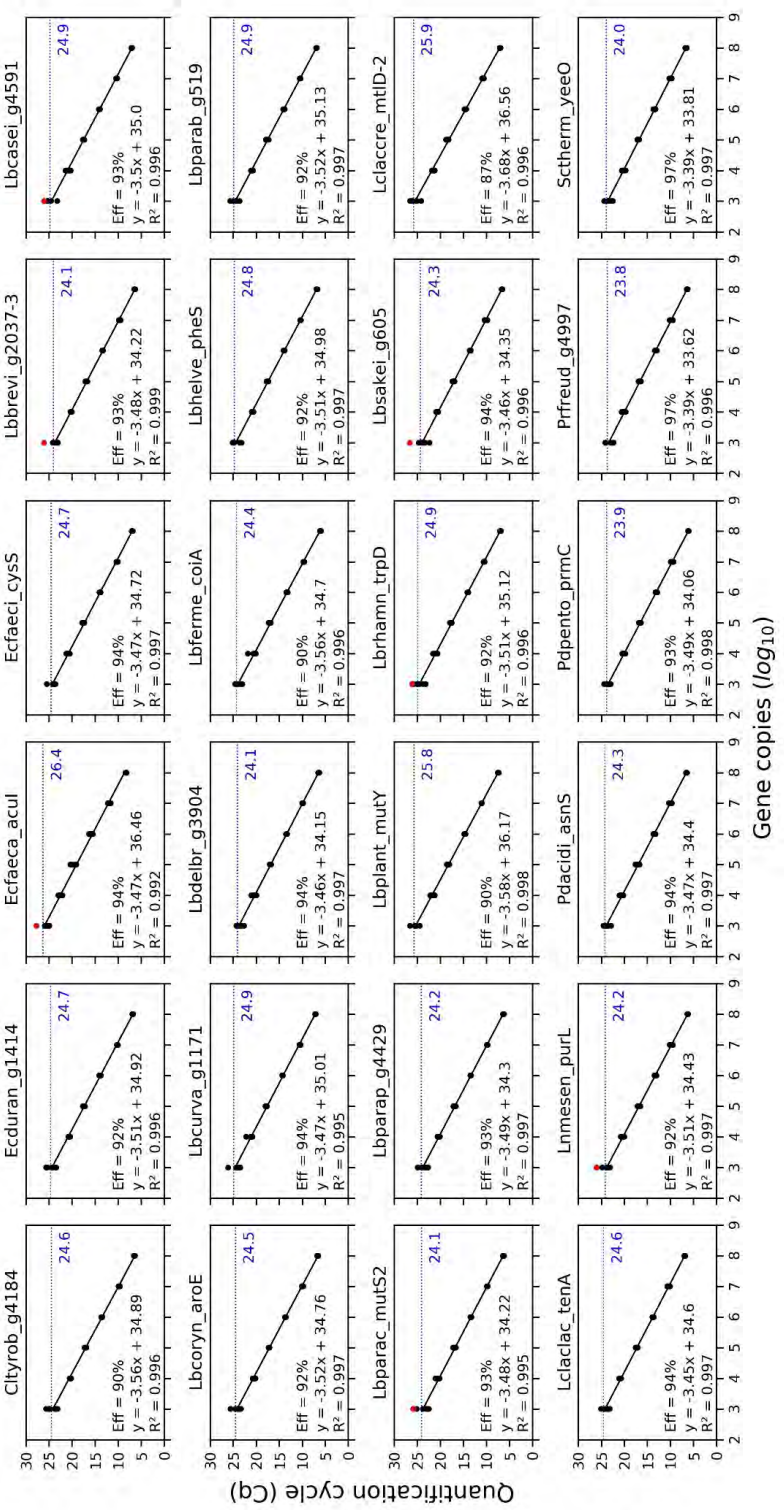


Figure S3.1: Standard calibration curves from tenfold dilution series of the HT-qPCR standard.
Logarithmic copies/μl are plotted against the quantification cycles (Cq) of the species-specific primer pairs. Red dots represent data points not considered for the linear regression calculation, due to a low quality score (>0.5). The efficiency, linear regression equation, and correlation coefficients (R^2) are shown in the lower left corner for each assay. The individual cut-off Cq values (corresponding to 800 copies/μl) are shown as a blue dotted line.

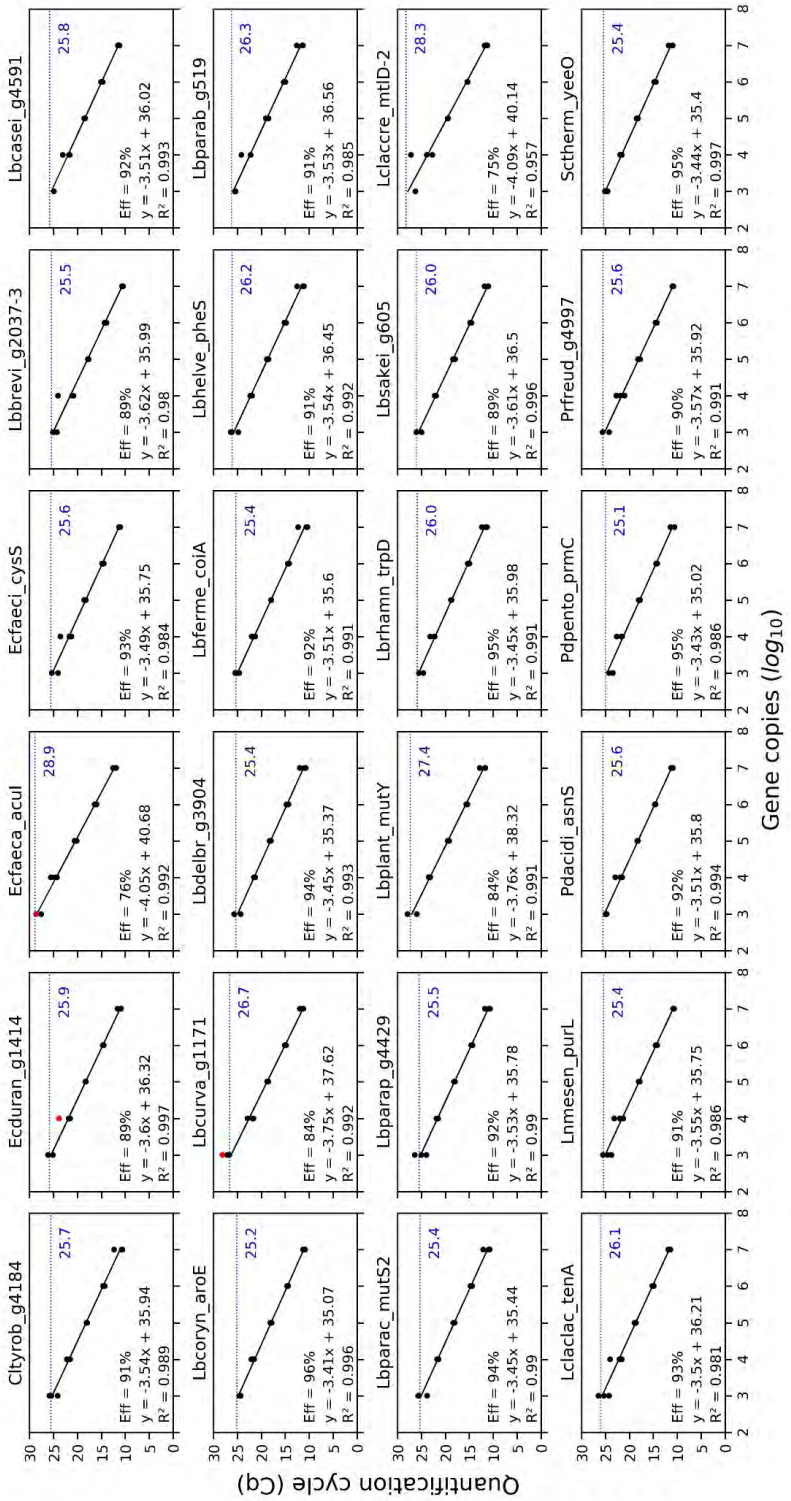


Figure S3.2: Standard calibration curves obtained from diluted series of the HT-qPCR standard used for quantification of bacterial target sequences in cheese samples.
Logarithmic copies/μl are plotted against the quantification cycles (Cq) of the species-specific primer pairs. Red dots represent data points not considered for the linear regression calculation, due to a low quality score (>0.5). The efficiency, linear regression equation, and correlation coefficients (R^2) are shown in the lower left corner for each assay. The individual cut-off Cq values (corresponding to 800 copies/μl) are shown as a blue dotted line.

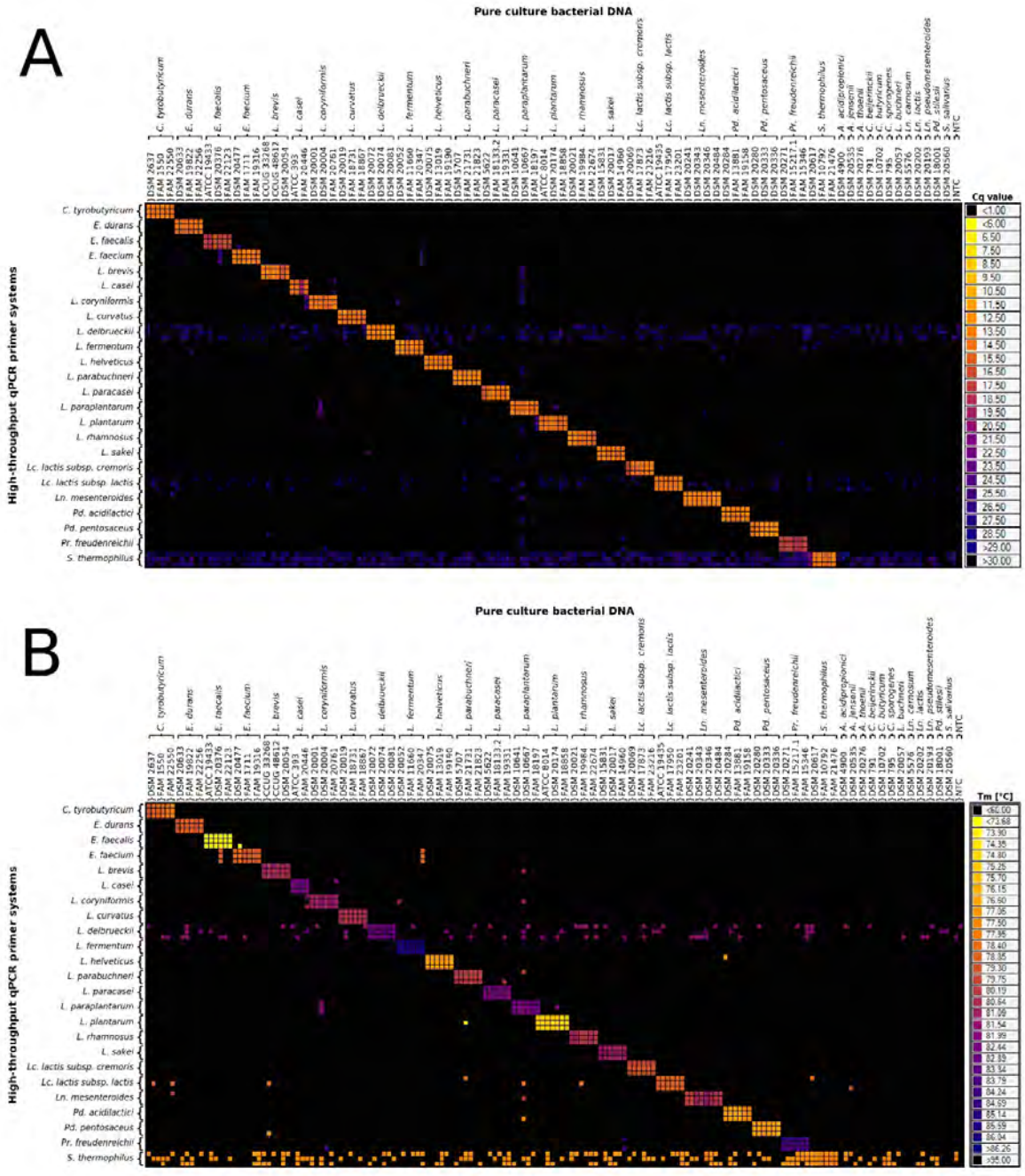


Figure S3.3: Heatmap of raw quantification cycle and melting curve analysis data.
(A) Raw Cq heatmap generated by the Fluidigm Real-Time PCR Analysis Software. Data points that did not reach the quality threshold (0.5) or the melting curve peak threshold (0.05) are marked with an X in the heatmap. (B) Raw T_m (inside peak detection range) heatmap generated by the Fluidigm Real-Time PCR Analysis Software. Abbreviations: *A.*, *Acidipropionibacterium*; *C.*, *Clostridium*; *E.*, *Enterococcus*; *L.*, *Lactobacillus*; *Lc.*, *Lactococcus*; *Ln.*, *Leuconostoc*; *Pd.*, *Pediococcus*; *Pr., Propionibacterium*; *S.*, *Streptococcus*; NTC, no template control.

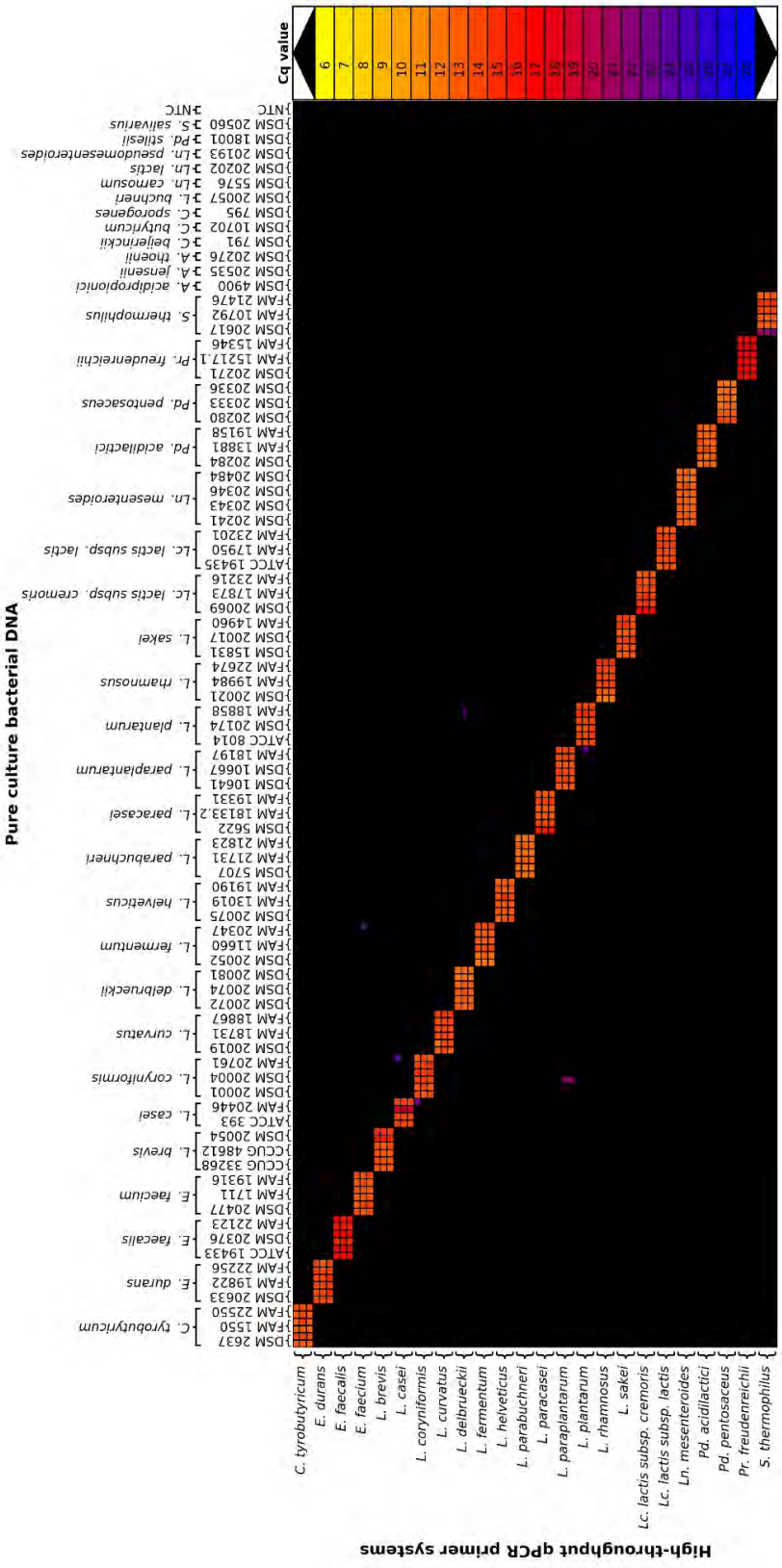


Figure S3.4: Heatmap of filtered quantification cycle (Cq) data.
Data cleaning was performed using the biomarkdataparser.py script. All flagged data points and data points with a Cq above the qPCR assay specific cut-off value were removed. Abbreviations: *A.*, *Acidipropionibacterium*; *C.*, *Clostridium*; *E.*, *Enterococcus*; *L.*, *Lactobacillus*; *Lc.*, *Lactococcus*; *Ln.*, *Leuconostoc*; *Pd.*, *Pediococcus*; *Pr.*, *Propionibacterium*; *S.*, *Streptococcus*; NTC, no template control.

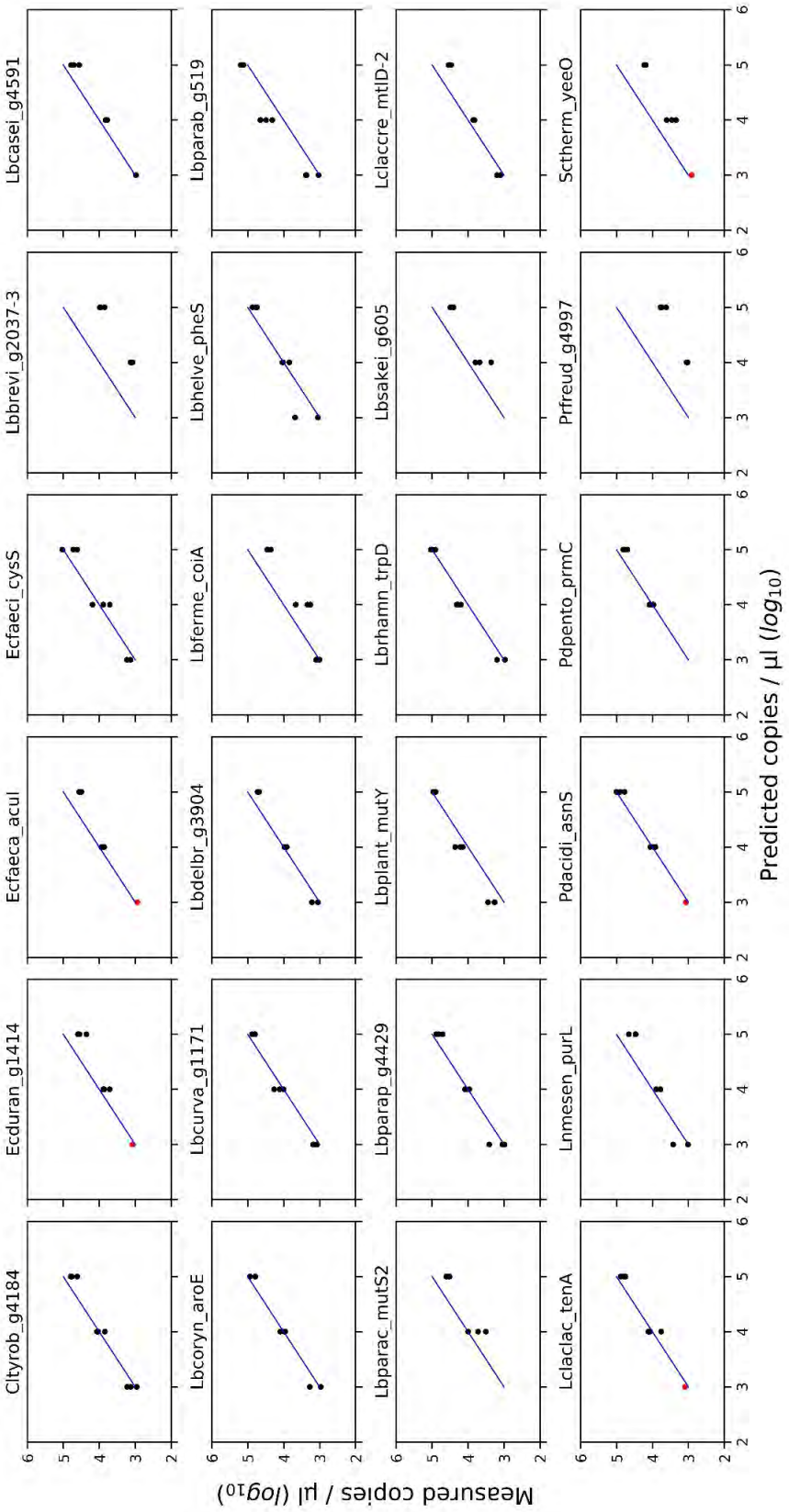


Figure S3.5: Plots of the predicted and measured population density of target species in diluted mock community samples without preamplification.
Predicted logarithmic copies/μl were plotted against the measured logarithmic copies/μl. The blue line indicates perfect correlation. Red points represent single data points not considered in the average data in **Figure 3.3**.

Supplementary data sheet 4

Raw data and scripts for data analysis

Supplementary data sheet 4 is a ZIP archive containing multiple files with different types of data, it is also available at Github https://github.com/biologger/htqpcr_validation_data. The data set is too large to include it in the thesis. A short description of the content is given below.

The data set contains raw data in tabular format (Assayvalidation_triplicates_rawdata.csv, CheeseSamples_validation_rawdata.csv), the scripts used for the data analysis (biomarkdataparser.py) and the Jupyter notebooks used for data analysis and preparation of the figures (bm_rawdata_annotation.ipynb, htqpcr_validation_figures.ipynb).

Supplementary data sheet 5

Raw data of the inclusivity assessment by standard qPCR experiments

Supplementary data sheet 5 is a large Excel workbook with tabular data, it is too large to include it in the thesis. A short description of the content is given below.

Raw qPCR data for the validation of the inclusivity ot the primer pairs Cltyrob_g4184, Ecduran_g1414, Ecfaeca_acuI, Ecfaeci_cysS, Lbbrevi_g2037-3, Lbcasei_g4591, Lbcoryn_aroE, Lbcurva_g1171, Lbdelbr_g3904, Lbferme_coiA, Lbhelve_pheS, Lbparab_g519, Lbparac_mutS2, Lbparap_g4429, Lbplant_mutY, Lbrhamn_trpD, Lbsakei_g605, Lclaccre_mtlD-2, Lclaclac_tenA, Lnmesen_purL, Pdacidi_asnS, Pdpento_prmC, Prfreud_g4997 and Sctherm_yeeO. The quantification cycle values and melting temperatures of the melt analysis are given for all target strains.

textbf

Supplementary data sheet 6

Identification of the contaminant of the bacterial culture stock (FAM 20347)

16S rRNA gene amplification and sequencing

A part of the 16S rRNA gene was amplified with the primers 16SUNI-L and 16SUNI-R (**Table SM3.1**) from the gDNA of the two isolates with different colony morphologies. The amplification was performed at a final volume of 25 ml, containing 200 nM of each primer, 2 U AmpliTaq Gold® Polymerase (Thermo Fisher Scientific, Waltham, MA, USA) 0.2 mM dNTPs, and 2.5 mL 10x PCR buffer I (Thermo Fisher Scientific, Waltham, MA, USA). The reactions were run on a TGradient thermocycler (Biometra, Göttingen, Germany) at 95 °C for 10 min followed by 35 cycles of 95 °C for 30 s, 54°C for 30 c, and 72 °C for 30 s, with a final extension for 7 min at 72 °C. The amplicon was purified using the MinElute PCR Purification kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. The purified PCR products were sequenced by Sanger sequencing using the primer 16SUNI-L and 16SUNI-R, respectively (Fasteris SA, Plan-les Ouates, Switzerland).

Table SM3.1: 16S rRNA gene primer used for amplification and sequencing

Primer	Sequence (5'-3')	Reference
16SUNI-L	AGA GTT TGA TCA TGG CTC AG	(Kuhnert et al., 1996)
16SUNI-R	GTG TGA CGG GCG GTG TGT AC	

BLAST search

An online BLAST search against the NCBI prokaryotic 16S ribosomal RNA database was performed with the sequences from Sanger sequencing.

Results

The BLAST results of the search against the 16S ribosomal RNA database have shown that the PCR product of the gDNA from colony type 1 shows the highest similarity to the 16S rRNA sequence of *Lactobacillus fermentum* strain CIP 102980 and NBRC 15885 (**Table SM3.2**). While the sequence of the second colony type shows the highest similarity to the 16S rRNA sequence of the *Enterococcus faecium* strains NBRC 100486, NBRC 100485 and ATCC 19434 (**Table SM3.3**). In summary, the contaminant of the *Lactobacillus fermentum* strain FAM 20347 stock was identified as an *Enterococcus faecium* strain.

Table SM3.2: BLAST results for the Query FAM20347.1_colony_type_1

QUERY	FAM20347.1_colony_type_1				
BLAST HIT	**Species**	Query coverage [%]	Identity [%]	Sequence title	Accession
1	***Lactobacillus fermentum***	99	99	Lactobacillus fermentum strain CIP 102980 16S ribosomal RNA, partial sequence	NR_104927.1
2	***Lactobacillus fermentum***	99	99	Lactobacillus fermentum strain NBRC 15885 16S ribosomal RNA gene, partial sequence	NR_113335.1
3	***Lactobacillus gorillae***	99	98	Lactobacillus gorillae strain KZ01 16S ribosomal RNA, partial sequence	NR_134066.1
4	***Lactobacillus ingluviei***	99	96	Lactobacillus ingluviei strain KR3 16S ribosomal RNA gene, partial sequence	NR_028810.1
5	***Lactobacillus gastricus***	99	95	Lactobacillus gastricus strain Kx156A7 16S ribosomal RNA, partial sequence	NR_029084.1

Table SM3.3: BLAST results for the Query FAM20347.1_colony_type_2

QUERY	FAM20347.1_colony_type_2				
BLAST HIT	**Species**	Query coverage [%]	Identity [%]	Sequence title	Accession
1	***Enterococcus faecium***	100	99	Enterococcus faecium strain NBRC 100486 16S ribosomal RNA gene, partial sequence	NR_113904.1
2	***Enterococcus faecium***	100	99	Enterococcus faecium strain NBRC 100485 16S ribosomal RNA gene, partial sequence	NR_113903.1
3	***Enterococcus faecium***	100	99	Enterococcus faecium strain ATCC 19434 16S ribosomal RNA gene, partial sequence	NR_115764.1
4	***Enterococcus faecium***	100	99	Enterococcus faecium strain DSM 20477 16S ribosomal RNA, partial sequence	NR_114742.1
5	***Enterococcus lactis***	99	99	Enterococcus lactis strain BT159 16S ribosomal RNA gene, partial sequence	NR_117562.1

References

Kuhnert, P., Capaul, S.E., Nicolet, J., and Frey, J. (1996). Phylogenetic positions of *Clostridium chauvoei* and *Clostridium septicum* based on 16S rRNA gene sequences. *Int J Syst Bacteriol* 46(4)**,** 1174-1176. doi: 10.1099/00207713-46-4-1174.

Chapter 4

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

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], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1], Pilar Junier[2]
[1]Agroscope, Schwarzenburgstrasse 161, CH-3003 Bern, Switzerland
[2]Laboratory of Microbiology, University of Neuchâtel, Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

Corresponding author:
Matthias Dreier
Email address: matthias.dreier@agroscope.admin.ch

4.1 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.

4.2 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 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.

4.3 Materials and Methods

4.3.1 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 PDO 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].

4.3.2 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×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).

4.3.3 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.

4.3.4 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.

4.3.5 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.

4.3.6 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.

4.3.7 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 AT**G CWG CCT CCC GTA GGA GT** -3'), and 45 µl Platinum™ PCR SuperMix High Fidelity (Thermo Fisher Scientific, Waltham, MA, USA). The 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).

4.3.8 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]).

4.3.9 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 the Github repository [49] (see also Supplementary Material, section 4.8).

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.

4.3.10 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.

4.3.11 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].

4.3.12 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 [53]. The best-scoring maximum likelihood tree was visualized using iTOL [54].

4.4 Results and Discussion

4.4.1 HT-qPCR and 16S rRNA gene amplicon sequencing results

The average sequencing depth for 16S rRNA gene amplicon sequencing was 471184 reads (range: 361496–632269). In total, 9894860 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* (**Figure 4.1**).

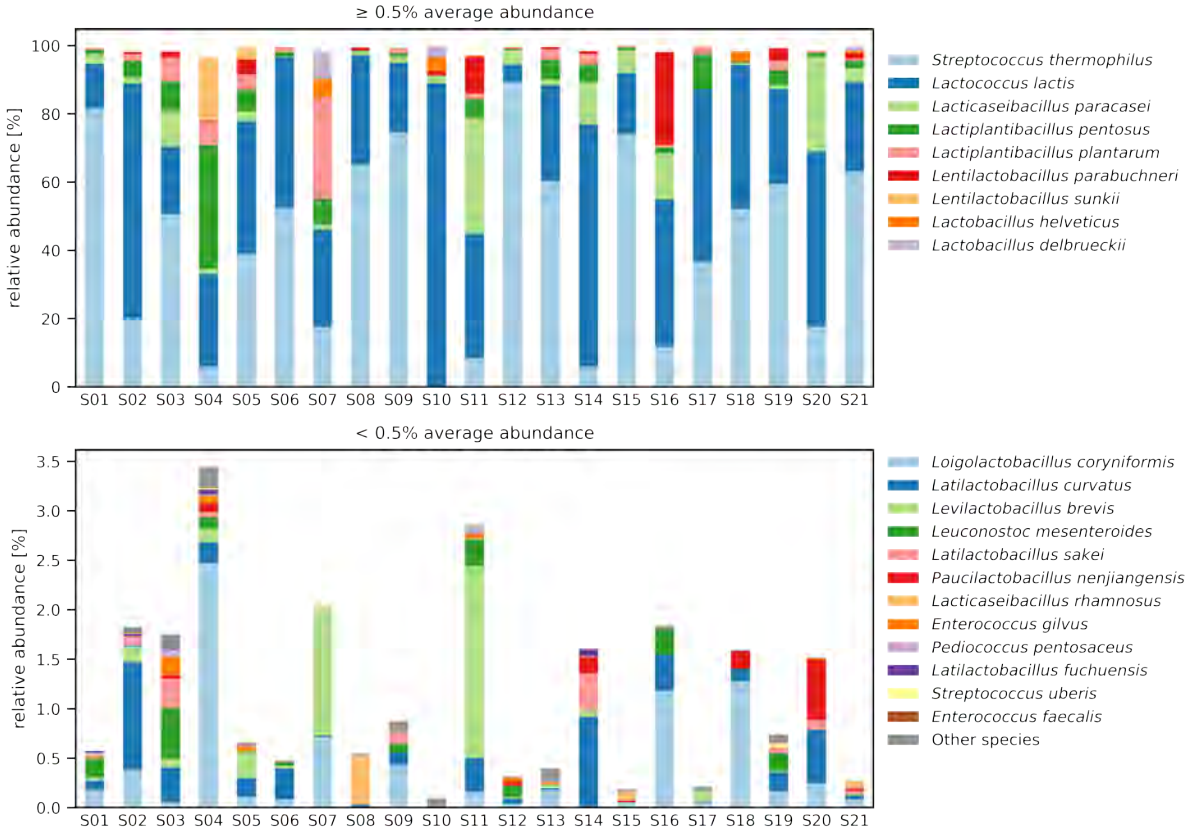


Figure 4.1: Bacterial community composition determined by 16S rRNA gene amplicon sequencing
Species detected in Raclette du Valais PDO cheese DNA samples (S01-S21) with an average relative abundance above 0.5% are shown in the upper panel and the species with a lower average abundance in the lower panel. 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.

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 (**Figure 4.2**).

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

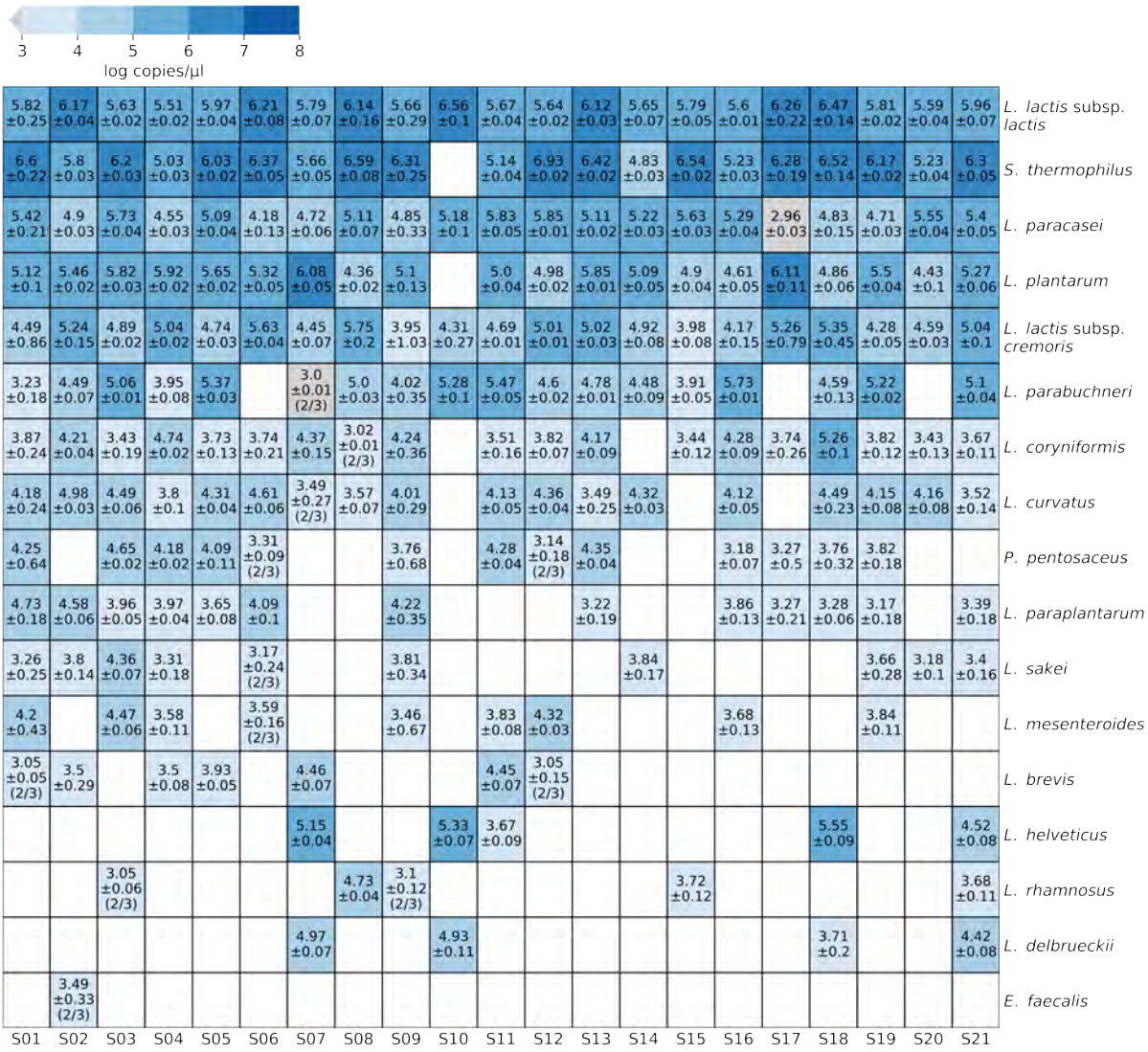


Figure 4.2: Heatmap of HT-qPCR results
The heatmap annotation depicts the average log copies/µl and the standard deviation of technical triplicates. When not all samples were positive, the number of positive samples out of the total number of samples is given in brackets.

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]).

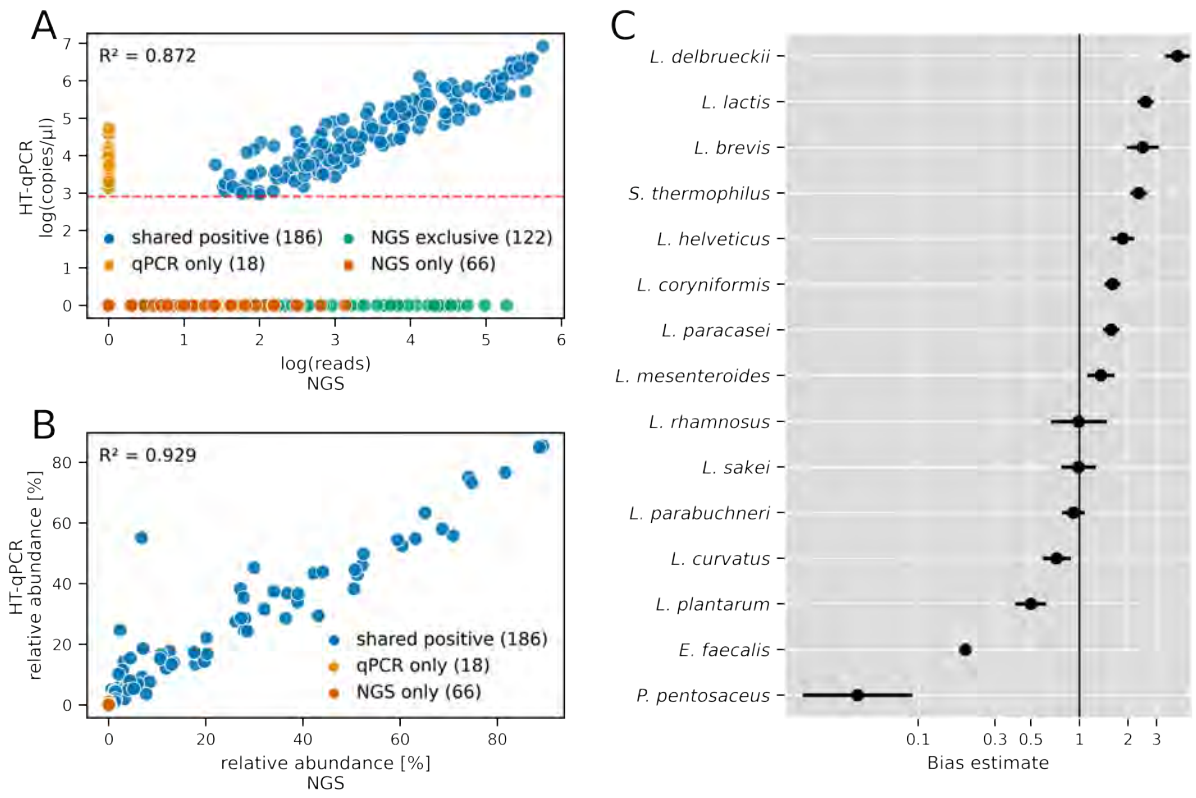


Figure 4.3: Comparison of HT-qPCR and 16S rRNA gene amplicon sequencing (NGS)
A) Logarithmic HT-qPCR data (y-axis) and logarithmic count data corresponding to the NGS approach (x-axis). The red line depicts the threshold of 800 copies/µl used for the HT-qPCR data analysis. **Shared positive**: Measurements 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 only 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.

4.4.2 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 (**Figure 4.3A**). 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 **Figure 4.3A** 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 (**Figure 4.3A**) and the relative abundance data (**Figure 4.3B**) 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 **Figure 4.3B** 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 4.1**. 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 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 53306 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 (**Supplementary Table S4.1**). 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 4.2**), 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 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].

4.4.3 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 (**Figure**

Table 4.1: 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.
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.

	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

4.3A 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 (**Figure 4.3C**). 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*.

4.4.4 Identification of possible causes for negative bias

Potential biases can be introduced at every step of the NGS protocols, from nucleic acid extraction, library 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 *Lpb. paraplantarum* (**Supplementary Figure S4.1A**). Multiple copies of the 16S rRNA gene in the representative genomes of *L. plantarum* and *L. paraplantarum* were not identical and contained single nucleotide

Table 4.2: Relative abundance data of species detected by 16S rRNA gene amplicon sequencing.
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.

	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	99.50	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	

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 Figure S4.1B**). The references for *L. plantarum* and *L. pentosus* contained only a single nucleotide transition. In the phylogenetic tree (**Supplementary Figure S4.1C**) 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*, *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 Figure S4.2**). 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 Figure S4.3**). 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; 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.

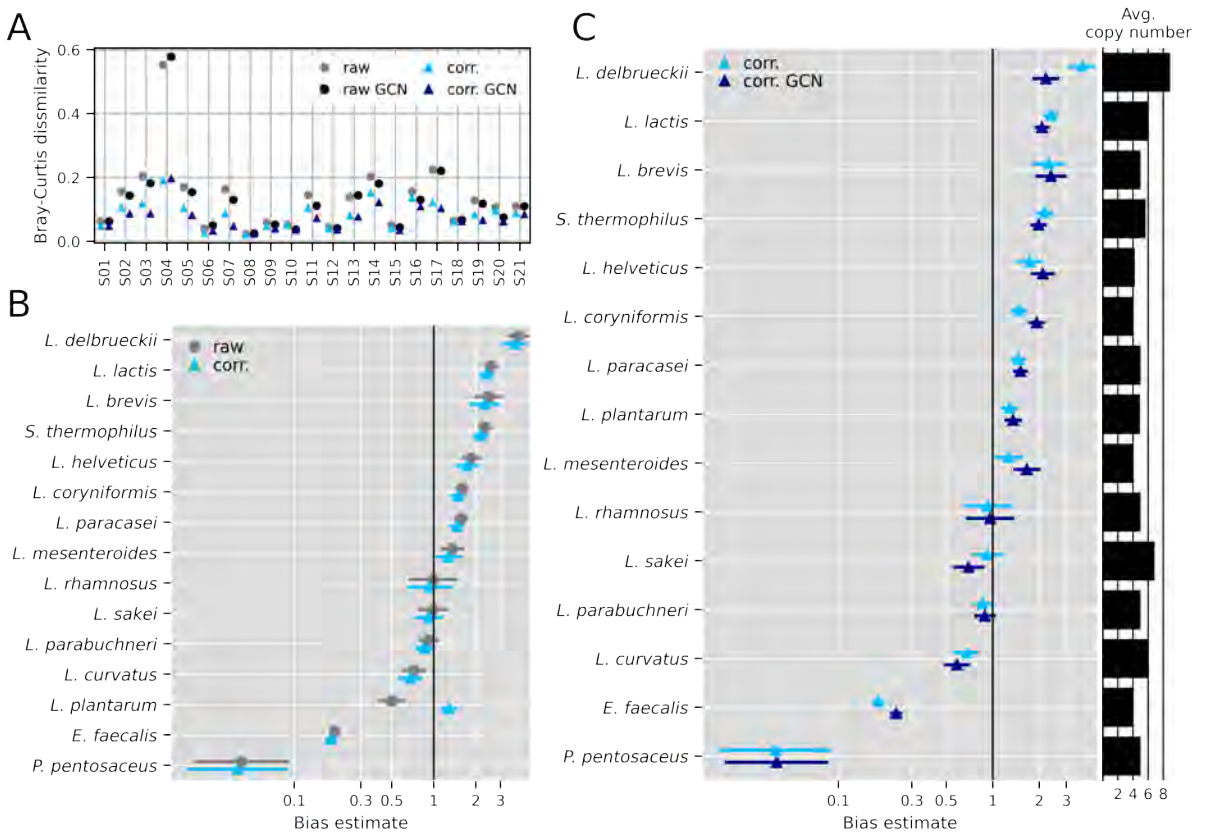


Figure 4.4: Bias estimates for raw and corrected data sets
A) Bray-Curtis dissimilarity of relative abundance data of 16S rRNA gene amplicon sequencing (NGS) to HT-qPCR relative abundance data (reference). Depicted are the dissimilarities for the raw and the corrected *Lactiplantibacillus plantarum* 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 and without 16S rRNA gene copy number correction.

4.4.5 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* (**Supplementary Table S4.2**).

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 (**Figure 4.4A**). 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 (**Figure 4.4B**). 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 (**Figure 4.4C**). 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].

4.4.6 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 (**Figure 4.5**). 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 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

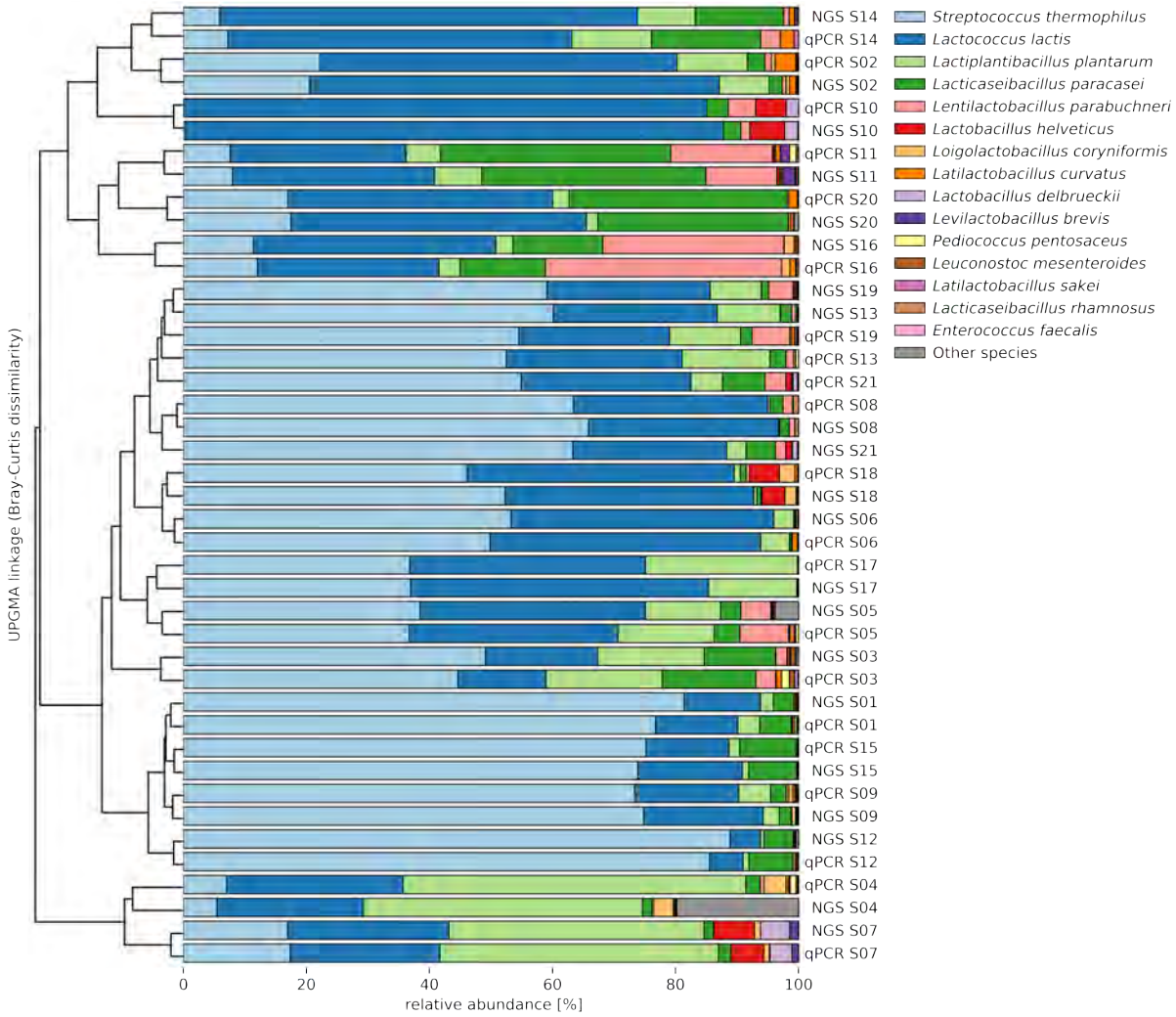


Figure 4.5: Comparison of HT-qPCR and 16S rRNA gene amplicon sequencing (NGS) data with corrections for *Lactiplantibacillus plantarum* group species and 16S rRNA gene copy normalization
Relative species compositions of the samples measured by HT-qPCR and NGS The samples are sorted and clustered according 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.

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, 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.

4.5 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.

4.6 Declarations

4.6.1 Ethics approval and consent to participate

Not applicable

4.6.2 Consent for publication

Not applicable

4.6.3 Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

4.6.4 Competing interests

The authors declare that they have no competing interests.

4.6.5 Funding

The authors received no external funding for this work.

4.6.6 Authors' contributions

MD, MM, HB, NS, DW, and PJ conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft. HB performed the NGS experiments; MM analyzed the data; and MD performed the HT-qPCR experiments, analyzed the data, and prepared the figures and tables.

4.6.7 Acknowledgements

We thank Daniel Marzohl, Nadine Sidler, Elvira Wagner, and Kotchanoot Srikham for their valuable technical help. The HT-qPCR data produced and analyzed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich. We especially thank Dr. Aria Minder and Silvia Kobel for their technical support in conducting the analyses.

4.7 References

1. Bharti R, Grimm DG. Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform. 2021;22(1):178-93; doi: 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; doi: 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); doi: 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; doi: 10.1093/nar/gkq873.
5. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ. 2018;6:e4652; doi: 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; doi: 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; doi: 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); doi: 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; doi: 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; doi: 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); doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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); doi: 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: 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; doi: 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; doi: 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; doi: 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; doi: 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 food-related microbiome studies. Front Microbiol. 2017;8:1829; doi: 10.3389/fmicb.2017.01829.
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; doi: 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); doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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(Database issue):D593-8; doi: 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. Computing in Science & Engineering. 2007;9(3):21-9; doi: 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; doi: 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. http://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. http://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; doi: 10.1038/s41592-019-0686-2.
46. Hunter JD. Matplotlib: a 2D graphics environment. Computing in Science & Engineering. 2007;9(3):90-5; doi: 10.1109/MCSE.2007.55.
47. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. 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: 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. http://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 in Molecular Biology (Clifton, NJ). 2014;1079:155-70; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 10.1371/journal.pone.0089680.
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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 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; doi: 10.7717/peerj.8544.

4.8 Supplementary Information

Data set S1

The data set used for this study is available at Github and Zenodo.
https://github.com/biologger/htqpcr_ngs_data
http://doi.org/10.5281/zenodo.4813080

The data set contains raw data in tabular format (V18-22-21_ASV_counts_table.csv, V18-22-21_htqpcr_rawdata.csv) and Jupyter notebook used for data analysis (htqpcr_ngs_comparison_R.ipynb).

Supplementary tables

Table S4.1: 16S rRNA gene copy numbers for taxa detected in Raclette du Valais PDO cheeses.
Source: https://rrndb.umms.med.umich.edu/static/download/rrnDB-5.7_pantaxa_stats_NCBI.tsv.zip

	Rank	N	Min.	Max.	Median	Avg.	SD
Aerococcus viridans	species	1	7	7	7	7	0
Clostridium tyrobutyricum	species	3	5	6	6	5.67	0.47
Companilactobacillus crustorum	species	1	4	4	4	4	0
Companilactobacillus farciminis	species	1	4	4	4	4	0
Corynebacterium glutamicum	species	30	5	6	6	5.93	0.25
Enterococcus faecalis	species	50	4	4	4	4	0
Enterococcus faecium	species	182	6	7	6	6.01	0.1
Enterococcus gilvus	species	1	6	6	6	6	0
Lacticaseibacillus paracasei	species	44	5	5	5	5	0
Lacticaseibacillus rhamnosus	species	26	5	5	5	5	0
Lactiplantibacillus pentosus	species	5	5	5	5	5	0
Lactiplantibacillus plantarum	species	135	1	5	5	4.93	0.5
Lactobacillus delbrueckii	species	25	8	11	9	8.92	0.63
Lactobacillus helveticus	species	17	4	5	4	4.24	0.42
Lactococcus lactis	species	48	5	6	6	5.94	0.24
Leuconostoc carnosum	species	2	4	4	4	4	0
Leuconostoc mesenteroides	species	20	3	4	4	3.95	0.22
Leuconostoc pseudomesenteroides	species	3	4	4	4	4	0
Levilactobacillus brevis	species	18	4	6	5	5	0.33
Pediococcus pentosaceus	species	18	5	5	5	5	0
Shigella dysenteriae	species	11	7	7	7	7	0
Staphylococcus aureus	species	575	1	8	6	5.58	0.79
Staphylococcus haemolyticus	species	10	1	7	6	5.5	1.57
Streptococcus dysgalactiae	species	24	5	8	6	5.92	0.64
Streptococcus gallolyticus	species	7	5	7	6	6	0.53
Streptococcus thermophilus	species	66	5	6	6	5.67	0.47
Streptococcus uberis	species	4	5	6	5.5	5.5	0.5
Turicibacter sanguinis	species	1	8	8	8	8	0
Weissella hellenica	species	2	8	8	8	8	0
Weissella paramesenteroides	species	1	8	8	8	8	0
Lentilactobacillus buchneri	species	3	5	5	5	5	0
Loigolactobacillus coryniformis	species	4	1	5	5	4	1.73
Latilactobacillus curvatus	species	12	6	6	6	6	0
Schleiferilactobacillus harbinensis	species	2	5	5	5	5	0
Paucilactobacillus nenjiangensis	species	1	5	5	5	5	0
Lentilactobacillus parabuchneri	species	3	5	5	5	5	0
Lactiplantibacillus paraplantarum	species	3	1	5	5	3.67	1.89
Latilactobacillus sakei	species	20	5	8	7	6.85	0.65
Companilactobacillus	genus	8	3	4		3.88	0.33
Enterococcus	genus	19	4	7		5.49	0.74
Lactobacillus	genus	28	3	11		4.78	1.06
Latilactobacillus	genus	3	5	8		6.28	0.4
Lentilactobacillus	genus	5	5	5		5	0
Loigolactobacillus	genus	2	1	5		4.5	0.5
Pediococcus	genus	5	4	5		4.4	0.49
Ruminococcaceae	family	6	2	6		3.5	1.29
Staphylococcus	genus	46	1	8		5.95	0.47
Streptococcaceae	family	2	1	9		4.98	0
Intestinibacter	genus	NA	NA	NA	NA	NA	NA

Table S4.2: Relative abundance data of species detected by 16S rRNA gene amplicon sequencing after *L. plantarum* group correction.
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, *L. pentosus* and *L. sunkii,* with the largest overall abundance are shown separately.

	Shared species		**Other species**		***L. sunkii***
	[%]	[n]	[%]	[n]	[%]
S01	99.96	13	0.03	3	0.01
S02	99.89	12	0.09	6	0.02
S03	99.64	14	0.36	6	
S04	80.96	14	0.44	11	18.6
S05	96.57	13	0.07	4	3.36
S06	99.99	12	0.01	5	
S07	99.99	12	0.01	1	
S08	99.98	10	0.02	3	
S09	99.89	12	0.11	3	
S10	99.91	7	0.09	1	
S11	99.95	15	0.05	3	
S12	99.64	13	0.09	5	0.27
S13	99.86	12	0.14	4	
S14	99.76	13	0.23	5	
S15	99.97	10	0.03	5	
S16	99.97	11	0.02	4	
S17	99.96	11	0.04	2	
S18	99.82	11	0.18	3	
S19	99.87	14	0.13	6	
S20	99.36	10	0.63	5	0.01
S21	99.97	13	0.03	7	

Supplementary figures

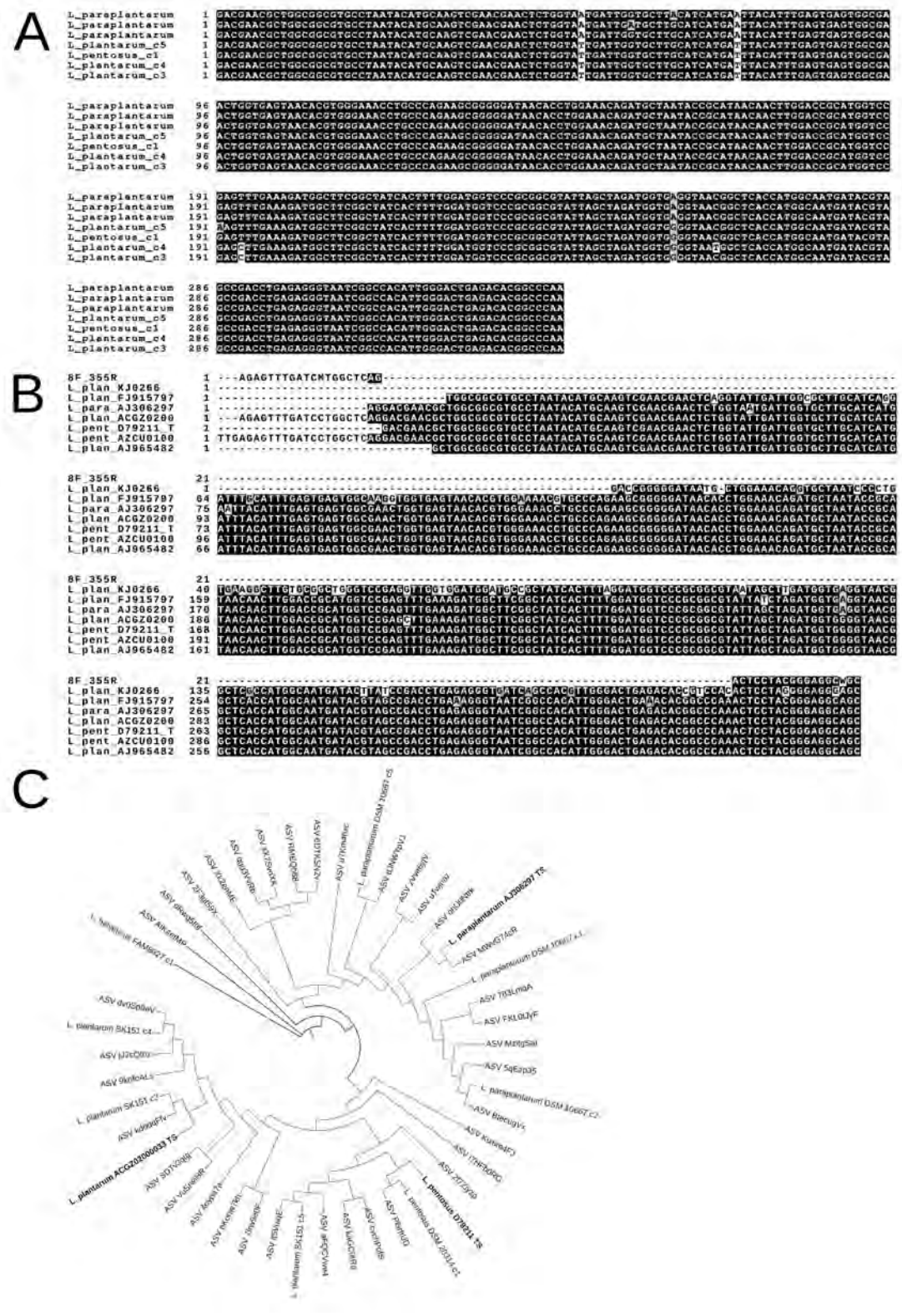


Figure S4.1: Analysis of the potential to differentiate *L. plantarum* group species based on V1–V2 variable regions of the 16S rRNA gene

A) Alignment of the V1–V2 16S rRNA gene region of the representative genomes of *L. plantarum*, *L. paraplantarum* and *L. pentosus*. B) Alignment of the DAIRYdb 16S rRNA gene reference sequences. C) The best-scoring maximum likelihood tree from alignments of the V1–V2 16S rRNA gene sequences including the representative genomes, the DAIRYdb reference sequences, and the ASVs for *L. plantarum*, *L. paraplantarum* and *L. pentosus*.

Figure S4.2: *L. plantarum* group discrimination of 16S rRNA gene amplicon sequencing targeting the V3–V4 region

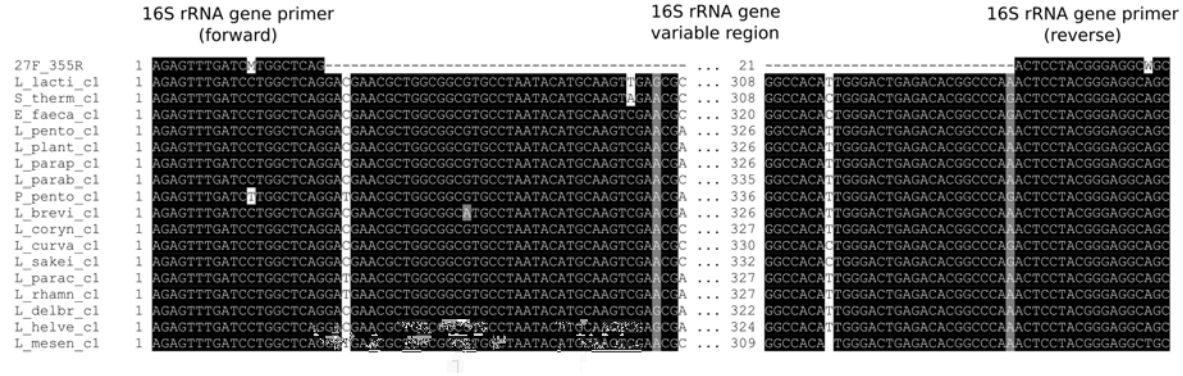


Figure S4.3: Alignment of the 16S rRNA gene primer binding regions

Chapter 5

Synthesis

5.1 General discussion

In the second chapter, we described the primer design pipeline, SpeciesPrimer, and designed and validated primers for four different species for qPCR assays. With the development of the SpeciesPrimer pipeline, which allows for a fully automated design of species-specific primers for qPCR, we created an important basis for finding optimal qPCR assays for different bacterial species in cheese. We required an automated approach for primer design because the requirements for high-throughput qPCR (HT-qPCR) are more demanding than for standard qPCR. For example, the specificity of a primer pair cannot be optimized by adjusting the annealing temperature, since all reactions must have exactly the same conditions. Therefore, the development and application of SpeciesPrimer provided the starting point for developing a modular HT-qPCR system.

The qPCR assays we developed with SpeciesPrimer use species-specific single-copy gene sequences as target sequences. In contrast, many older qPCR assays used in food microbiology used universal genes, such as the 16S rRNA gene, *tuf*, or *pheS*, as target sequences (Postollec et al., 2011). The ever-increasing availability of sequenced genomes has allowed us to identify specific sequences for each of the 23 targeted bacterial species/subspecies. This had the advantage that more candidate target sequences and primers were available that we could test and optimize our choice for the given annealing temperature, since we were not limited to the variable regions of universal genes.

For the implementation of SpeciesPrimer, we relied on Docker to provide a container with all dependencies that worked on different platforms and operating systems. Software environments with defined versions or version control of all dependencies are important for the reproducibility of analyses. Data scientists and bioinformaticians are therefore increasingly using virtual environments and container solutions, such as Docker (Bordin and Devos, 2018; Menegidio et al., 2018; Panda et al., 2020), Anaconda and Bioconda (Karunanithi et al., 2019; Rautiainen and Marschall, 2020; Singh and Wurtele, 2021), and web-based platforms, such as Galaxy (Afgan et al., 2018; McGowan et al., 2020; Wibowo et al., 2021). Through dockerization, even deprecated dependencies can still be used and enable the continued use of the SpeciesPrimer pipeline by the broader scientific community. Moreover, customizing and extending its functionalities is also possible and could further improve the tool's utility.

Some extensions of the functionality for SpeciesPrimer that have been requested by a few users from the scientific community include, for example, the feature to design strain-specific and pathovar-specific primers or the option to perform automated primer design for viruses. For the design of strain-or pathovar-specific primers, only the genes of the target group would need to be selected and checked for specificity instead of the core genes of the species, an adaptation that is non-trivial but feasible and that would not affect the subsequent steps of the pipeline. A similar pipeline could also be envisioned for viruses, although quality control would require an average nucleotide identity approach rather than a BLAST search with conserved universal genes. Further, there are far fewer candidate target genes available. Therefore, it is likely that an option to fine-tune a threshold for the similarity of non-target sequences would need to be implemented. Considering a few forks (a copy of the repository that other users can work on independently) on GitHub, it would also be encouraging if this project was continued and further developed with help from the scientific community.

The above-mentioned applications are relatively straightforward to implement because they use most of the selected tools of the SpeciesPrimer pipeline and only the workflow would need to be adapted. At the time of development, several factors were important for the selection of the tools, such as availability, ease of use, good documentation, and economical use of computer resources such as processor performance and random access memory. Given the rapid development of computer hardware and bioinformatics' software, other applications based on parts of the SpeciesPrimer pipeline may become possible. For example, new tools for pan-genome analysis could remove the species-level limitation. For example, it may become possible to select target sequences for primer design based on ANI values or for specific taxonomic levels, e.g. the genus level. It would also be conceivable to select strain-specific target sequences based on one-to-one comparisons between all input genomes, e.g., using k-mer-based algorithms. These resource-intensive applications could provide the target sequences, while the *in silico* primer validation is performed using SpeciesPrimer's established primer quality control method. Further, instead of automatically selecting the most conserved sequences, an option to select sequences with maximum variability among different input genomes for primer design could be useful for strain typing applications, such as high-resolution melting analyses or single- and multilocus sequence typing approaches.

In the third chapter, the development and validation of an HT-qPCR system for the study of the cheese microbiota are described and discussed. We have demonstrated the feasibility of HT-qPCR analysis in raw milk cheese, indicated potential applications for cheese research, and highlighted the advantages and limitations of this method. To our knowledge, the developed HT-qPCR system is the first quantitative culture-independent method able to simultaneously quantify 23 bacterial species often found in raw milk cheeses.

In the current state, the HT-qPCR system can be a useful tool for researchers to study LAB dynamics in starters, natural whey cultures, and the cheese core during ripening. Similar to what has already been done in several studies

of starter bacteria, such as *Streptococcus thermophilus* and *Lactococcus lactis*, or for NSLAB species for various types of cheese (Desfosses-Foucault et al., 2013; Ganesan et al., 2014; Pega et al., 2016; Ruggirello et al., 2016; Ruggirello et al., 2018; Bertani et al., 2020). The advantage of using HT-qPCR is that more species can be measured simultaneously. This facilitates the analysis of raw milk cheeses with low scalding temperatures and relatively high bacterial diversity.

To extend the usefulness of the analysis of raw milk cheeses, it would be desirable to integrate the missing bacterial species identified by comparison with NGS (especially *Lactiplantibacillus pentosus* and *Lentilactobacillus sunkii*) into the modular system. In addition, reference data for different cheese types should be collected, similar to the existing reference data for several chemical and biochemical parameters. In this way, microbial populations typical of different cheese types could be identified. This information could help define thresholds for certain species to check for product quality or to identify counterfeit cheeses, for example, in cases where certain starter cultures, a certain proportion of natural whey cultures, or the use of raw milk are required in the cheese manufacturers' guidelines.

Furthermore, many fermented foods benefit from similar bacterial communities like cheese; therefore, some overlap of LAB species between cheese and other fermented foods, such as kimchi, sauerkraut, sausages, and fermented fish, can be observed (Plengvidhya et al., 2007; Cocolin et al., 2009; Jung et al., 2011). In addition, also several fermented beverages, not only of milk origin, contain some common LAB species (Tamang et al., 2016). Therefore, to extend the application to other fermented foods, some qPCR assays could simply be exchanged for the specific food to be analyzed. Using the standardized primer design and validation process described in Chapters 2 and 3, we may further facilitate and accelerate the integration of new primer pairs into the HT-qPCR system. The influence of LAB and probiotics on the composition of the gut microbiota and the impact on health are areas of intense investigation, which could also have some interesting applications for the HT-qPCR system developed in the future.

In addition to species quantification, other target genes could be quantified with HT-qPCR, such as the histidine decarboxylase (*hdc*) or tyrosine decarboxylase (*tdc*) of different LAB species. This would be useful for assessing whether biogenic amine production is a species or a strain property, e.g., by differential analysis of species abundance and levels of the decarboxylase genes and for assessing the potential of biogenic amine production in the early stages of cheese ripening. However, for the final assessment of cheese quality, other parameters have to be considered that cannot be determined by HT-qPCR alone. The presence of decarboxylase genes does not necessarily predict the presence of biogenic amines, as other genes are involved in metabolic pathways, and the expression of these genes determines whether large amounts of biogenic amines are produced. Therefore, methods, such as high-performance liquid chromatography (HPLC), which measures the concentration of biogenic amines, remain essential for assessing food safety.

The inability to predict the accumulation of metabolites concerns not only HT-qPCR. The same applies to most microbiological or molecular biological methods used in complex and dynamic microbiological systems. However, merging microbial community composition data with metabolomic data is increasingly attempted, although it remains challenging to draw meaningful conclusions based on correlations between detection of taxonomic groups and metabolites.

In cheeses with faulty fermentations, there is a strong correlation between the detection of bacteria from the genus *Propionibacterium* and the species *Clostridium tyrobutyricum* and the accumulation of propionic acid and butyric acid, respectively (Klijn et al., 1995; Gómez-Torres et al., 2015; Turgay et al., 2018). A recent study found several correlations between different genera (*Streptococcus*, *Lactococcus*, *Lactobacillus*, *Leuconostoc*, *Pediococcus* and *Macrococcus*), and specific metabolites in the core and the surface of cheddar cheeses (Afshari et al., 2020). However, such findings are difficult to break down to the species level due to the high variability within these genera, especially *Lactobacillus*.

A recently published deep metagenomics and metabolomics approach has even shown a high variability in the genomes of different strains for the same species in cheese (Yang et al., 2021). The correlations for metabolites and different strains of the same species were largely inconsistent and even opposite trends were observed. The authors suggested that the flavor and cheese quality is rather determined by the individual strains and not on the species-level.

Since our newly developed HT-qPCR system allows only the quantification on the species-level, the identification of correlations with metabolites might prove to be difficult. The unambiguous identification of bacteria causing defects in cheese, apart from taxa with known causality, is very unlikely. However, since in practice samples with known defects are usually submitted for analysis, HT-qPCR can still serve as a rapid preliminary screening of known potentially harmful bacterial species and facilitate identification of underlying problems. For more accurate tracing and identification of problematic strains, typing approaches as well as deep

Table 5.1: Simplified estimation of the costs of reagents and consumables for the analysis of 56 cheese DNA samples with 24 assays, including standards and non-template controls, in technical triplicates. This corresponds to a full Dynamic Array 192×24 Gene Expression chip with 4608 reactions. Prices according to the Center for Genetic Diversity (GDC) price list (March 2020).

	Cost/unit	Unit	Amount		Costs [CHF]	
			Rotor-Gene	HT-qPCR	Rotor-Gene	HT-qPCR
qPCR Master Mix	0.08	µl	27648	649	2211.84	51.92
10 µl filter tips	0.07	tip	4608	216	305.76	14.33
Rotor-Gene tubes[1]	0.17	tube	4608	0	792.58	0.00
Mix plates	3.62	plate	0	3	0.00	10.86
192x24 GE chip	800	chip	0	1	0.00	800.00
				Total	**3310.18**	**877.11**
				Price/reaction	0.72	0.19
				Price/sample	59.11	15.66

[1]List price (LABGENE Scientific SA, 2021)

metagenomic sequencing combined with other omics approaches are potential solutions that could be used in the future.

HT-qPCR using target genes for specific metabolic pathways could also be of interest for other applications in the field of food research and food safety, for example, for toxin genes in *Staphylococcus aureus* or for the differentiation of *C. sporogenes* and *C. botulinum* based on genes of the botulinum neurotoxin gene cluster.

The long-term vision for developing this method was to use it as a diagnostic approach, for example, for the analysis of cheese defects or routine screenings to detect potential causes of microbially caused quality problems at an early stage. However, future use of the HT-qPCR system as a diagnostic tool for routine screenings in cheese dairies and the industry will require further validation. The repeatability and analytical sensitivity of the entire workflow, from DNA extraction to HT-qPCR analysis, would need to be validated. This could be accomplished with spiking experiments. For this, homogenized cheese samples could be spiked with a dilution series of bacterial strains with a known concentration, and the recovery could then be measured after the complete analysis pipeline. The same validation could be performed with bacterial species that are part of starters, for example, by heat treatment or sterilization of the homogenized cheese samples, to inactivate the starter bacteria but retain the components of the cheese matrix in the samples. Subsequently, by differentiating viable spiked bacteria from dead bacteria, for example, using propidium monoazide (PMA) treatment, recovery could be assessed. PMA treatment has already been successfully used in studies of cheddar cheese to distinguish viable bacteria from bacteria with damaged cell membranes using culture-independent molecular methods, such as qPCR (Desfosses-Foucault et al., 2012; Ganesan et al., 2014; Xue et al., 2021).

The previously described validation would allow the HT-qPCR method to be used for the application of early identification of possible microbial causes of cheese defects. The HT-qPCR system is faster than microbiological plate counting methods, and most of the chemical and biochemical methods used to date are at comparable costs for chemicals and laboratory consumables. A rough estimate of the costs of reagents and consumables for standard qPCR (Rotor-Gene 3000) and the HT-qPCR approach (Biomark) showed that HT-qPCR requires about 40 times less master mix and that the total costs of reagents and consumables are about four times lower (**Table 5.1**). In addition, with the HT-qPCR approach, 4608 qPCR reactions can be analyzed in half a day, whereas with the Rotor-Gene 3000 (72 reactions per run), the same number of reactions would require 64 runs, which is equivalent to at least eight days of work. The actual measurements of 56 cheese samples and 24 assays with the Rotor-Gene 3000 (72 reactions/run) would require even more reactions than estimated here, because the standards and non-template controls have to be included in each run. The costs of DNA extraction are the same for both approaches. DNA extraction from 56 cheese samples (for one HT-qPCR run at full capacity) takes one to two days, and the cost for the DNA extraction kit is about CHF 650 (~ CHF 12/extraction) according to current list prices (June 2021). However, detailed calculations for the cost of labor, reports, laboratories, and equipment would still need to be performed for HT-qPCR in cases where it will be used as a diagnostic method. For comparison, the cost for the quantitative determination of *Listeria monocytogenes* and the detection of *Salmonella* spp. in diagnostics is CHF 60; however, these diagnostic methods are ISO certified and the price includes analyses, processing costs, and VAT (Agroscope, 2020). The analysis of biogenic amines in cheese costs CHF 265, and the ultra-performance liquid chromatography (UPLC) method is rather time consuming (Ascone et al., 2017; Agroscope, 2020).

Compared to the chemical and biochemical analysis methods that have often been used to identify defects, the HT-qPCR approach has certain limitations. For example, the detection of *L. parabuchneri* correlates with

histamine accumulation in cheeses, but there are also strains that do not possess the *hdc* gene (Wüthrich et al., 2017). Therefore, for the quantification of the amount of biogenic amines formed, classical chemical methods are still required. However, HT-qPCR can provide an initial assessment of potential detrimental bacterial species before metabolites have accumulated and therefore provide an early evaluation of risks for potential cheese defects.

Another possible application is the rapid screening of raw milk cheeses for quality-relevant bacterial species to assess cheese quality in cheese dairies and ripening facilities. Especially cheese dairies that use raw milk for cheese production, including alpine cheese dairies, are interested in methods to quickly test the milk and cheese quality as cost efficient as possible (Turgay et al., 2018; Burtscher et al., 2020). The economic success of alpine farms is closely linked to the value added by the production of cheese and its marketing to the end consumer. The alpine dairies bear production, quality, and marketing risks, which means that cheese quality problems can have serious financial consequences (Lauber et al., 2014).

The HT-qPCR system significantly reduces the cost per qPCR reaction, although only if the capacity of the microfluidic chips is fully exploited. A further reduction in the cost of analyses is most likely possible by reducing labor, since the cost of laboratory consumables and chemicals has already been minimized using microfluidic chips. Thus, the labor required for chip preparation is already relatively low compared to conventional qPCR methods. A further reduction in labor would be possible using pipetting robots, but these would have to be optimized for the very small volumes, which is often not the case with the current, less expensive models. However, the greatest labor savings are possible with the automation of DNA extraction, which is currently the most time-consuming aspect. With the development of a fast and efficient automated method for DNA extraction, the prices for individual analyses could then be significantly reduced. Nevertheless, it is important to note that the initial purchase costs for the Biomark HD instrument, the integrated fluidic circuit controller, and robots for pipetting and DNA extraction are considerable. Moreover, the analyses can only be performed in well-equipped laboratories and not on-site.

Consequently, the HT-qPCR method will probably not be affordable for smaller cheese dairies and alpine farms in the near future. However, if these cheese dairies join forces to coordinate sample collection and transport, routine screening could also be conducted for small and alpine dairies. At least in the case of long-term problems with the quality of the cheese, the financial investment would be justified as the costs for months of maturing at the affineurs must be added to the costs for production. Prices for declassified cheeses are significantly lower, and they can only be sold for industrial processing into grated and processed cheeses.

A concrete application for the HT-qPCR system could be the continuous monitoring of young cheeses (aged for about two months) from all producers of a protected designation of origin (PDO) variety association, such as Emmental PDO (about 100 producers) or Gruyere PDO (about 155 producers). The PDO variety associations have strict guidelines on traditional production, which ensure that the quality and taste of their products stand out from industrially produced cheeses.

Monitoring microbial diversity over time and collecting reference data could help to quickly identify potential quality problems. Nowadays, "taxation", (quality grading of cheeses) is done after 3 months for these varieties, so valuable time is lost to detect quality defects and initiate measures to improve cheese quality. Thanks to the reduced analysis costs, HT-qPCR could enable a preventive quality control at an early stage of maturation.

The verification of the variety-specific composition of the cheese microbiota makes it also possible to detect fraud due to non-traditional production of cheese. For example, in the case of raw milk cheeses with low scalding temperatures, if the milk has been thermized or pasteurized to avoid potential microbiological defects, this would be noticeable due to the greatly reduced biodiversity of the microbiota.

Prolonged problems with cheese quality can lead to major losses. If the quality deficiencies are of microbial origin, the identification of the causes at the species level allows the initiation of targeted investigations that help to identify and eliminate the sources of contamination. Species identification can clarify whether the bacteria originate from the milk or from the production environment of the cheese dairy. If the milk is affected, the individual suppliers could examine the farm milk and clarify the cause with the respective producer. In the case of insufficient cleaning of the milking facilities, for example, qPCR analyses of samples from the piping systems and containers could identify possible sources of contamination. Approaches using species-specific qPCR to identify isolates and subsequent typing have already been successfully used to detect histamine-producing strains of *L. parabuchneri* (Ascone et al., 2017).

In the fourth chapter, we discussed the performance of the HT-qPCR approach for quantifying the dominant bacterial species in the raw milk cheese core and compared it with the performance of 16S rRNA gene amplicon sequencing. During the literature review and analysis of the data for this study, it was remarkable to learn that there are still many challenges in the appropriate analysis and interpretation of NGS data (Gloor et al., 2017; Leite and Kuramae, 2020). Although there are now numerous approaches to analyzing NGS data using compositional methods and mathematical models adjusted to the compositional nature of the data, many of these methods are quite complex. An often-recurring challenge in studies on microbiomes in ecology, medicine, or food microbiology

is differential abundance analysis, for example, between different sites or samples before and after a particular treatment. In medicine, differences in the microbial composition of body sites between individuals are used to study the association of microbial community composition with disease (Costello et al., 2009). Similarly, in food microbiology, the microbiomes from batches of cheese of good quality and batches with defects are compared to identify the microbial causes of cheese defects (Quigley et al., 2016; Ritschard et al., 2018). Bacteria that are exclusively found in cheeses with a certain defect have already been identified by amplicon sequencing (Quigley et al., 2016). Nonetheless, differential abundance analyses using amplicon sequencing remain challenging because relative abundances make it difficult to identify the taxa responsible for the observed effects when measured in both batches/groups. The following example illustrates this: In a simple sample with three different taxa, an increase in the relative abundance of taxon 1 is observed after a certain treatment. By comparing the relative composition before and after the treatment, it is not possible to determine beyond doubt whether the abundance of taxon 1 has increased or whether the abundance of taxon 2 and 3 has decreased, while the abundance of taxon 1 has not changed. In contrast, qPCR makes it very straightforward to precisely measure the abundance of taxon 1 before and after treatment by estimating the number of molecules in the sample compared to a known standard. Therefore, we see great potential for using NGS and HT-qPCR in combination to estimate the bias of NGS measurements for specific taxa to improve differential abundance analyses.

A perspective on how the latest achievements in microfluidics, HT-qPCR, and NGS could be used in combination for the development of a quantitative high-throughput sequence typing approach, e.g., for strain typing. With the growing number of prokaryotic genomes available, specific sequences of a certain taxonomic group could be found containing conserved sequences for the primers and variable regions to distinguish different sequence variants. Furthermore, there are already microfluidic chips whose PCR reactions can be further used for NGS after HT-qPCR. Thus, the quantitative qPCR data could be combined with the higher resolved information of the different amplicon sequence variants by NGS to obtain a strain-level resolution of the composition of the microbiota. An approach similarly combining HT-qPCR and NGS, however, without strain-level resolution, has been successfully used to study the intestinal microbiota from piglets (Hermann-Bank et al., 2013).

The bacterial community in cheese is shaped by harsh conditions during production and ripening. The resulting limited diversity makes cheese an interesting model system for studying complex microbial communities. Therefore, the methods developed, such as species-specific primer design, standardized protocols for validating HT-qPCR primer systems, and insights into identifying biases in 16S rRNA gene amplicon sequencing approaches, are not only relevant to cheese research but may also be useful for other ecosystems and diagnostic approaches. HT-qPCR could be interesting for the study of different microbiomes and for applications in medical diagnostics, with the aim of determining the absolute cell density of bacteria and not only their relative proportion. This despite recent progress in differential abundance analysis for NGS and third-generation sequencing approaches, as these approaches are still quite complex and not always successful. Absolute quantification by qPCR may be particularly advantageous when sampling depth varies widely and for samples of different origins, as well as in the velocity of obtaining results. Interesting approaches could include the development of high-throughput chips for the quantification of common viruses and other pathogens in hospitals or screening antibiotic resistance genes in food, livestock, water, and soil. HT-qPCR is already used in a number of applications, including typing viruses (Goecke et al., 2018), analyzing community dynamics in soil samples (Kleyer et al., 2017), and profiling functional genes of bacteria (Zhao et al., 2019).

The current limitation of our approach to the species-level could be overcome by extending the functionality of SpeciesPrimer or by the availability of similar primer design tools. New features for SpeciesPrimer, such as strain typing or strain-specific primer systems, could enable a broader range of applications. An advantage over other primer design tools is that primer parameters have already been optimized and validated for both standard qPCR and HT-qPCR. As new sequencing technologies and multi-omics approaches are rapidly developed, we will soon be able to reveal many new exciting discoveries in microbial ecology. These exploratory studies will certainly generate new questions and new targets for more detailed investigations, and likely, some of them could be investigated much faster, more economically, and more efficiently over a long time and with numerous samples using targeted approaches such as HT-qPCR.

5.2 Conclusions

The research presented in this thesis has allowed the development of an HT-qPCR method to quantify 23 bacterial species relevant to raw milk cheese quality. To our knowledge, this is the first qPCR approach capable of simultaneously quantifying so many different bacterial species commonly found in cheese.

In this thesis, we successfully designed 23 new primer systems with identical annealing temperatures that demonstrated high specificity and efficiency in classical qPCR and HT-qPCR. Automation of the primer design

process has made development more straightforward and efficient and has shortened the time-consuming validation process in the laboratory. Validation of the primer systems with standard qPCR and HT-qPCR has shown that the selected target genes can distinguish closely related species, such as the *Lacticaseibacillus casei* group species and two species of the *L. plantarum* group, some of which are poorly distinguished with primers for the partial 16S rRNA gene. The SpeciesPrimer pipeline could be used in other research areas besides fermented foods to find promising primers for PCR-based methods.

By applying the HT-qPCR system to inoculated model cheeses and commercial cheeses with quality problems, we demonstrated that the quantification of quality-relevant bacterial species could provide important information regarding the possible causes of cheese quality defects. However, for metabolic pathways that are strain-specific, no clear predictions can be made based on species detection; furthermore, HT-qPCR does not allow a prediction of the concentration of accumulated metabolites.

The comparison of HT-qPCR with NGS showed that a very high proportion of the most abundant species could be quantified using HT-qPCR, and that a future extension with only a few additional species would cover the most important and most abundant representatives of the bacterial community in the core of raw milk cheeses. The detailed comparison of the two methods made it possible to identify the bias present in the measurements of the NGS protocol used. These findings may help to improve the method, for example, in the choice of primers, and could be used to optimize the used reference database DAIRYdb. For the first time, HT-qPCR allowed us to perform a cross-method validation of the developed NGS workflow and allowed for a combined validation of the microbial community composition and the correct assignment at the species level.

To use HT-qPCR for diagnostics and monitor cheese quality in practice, the whole process, from DNA extraction to data analysis, still needs to be validated. In addition, the collection of reference data from different cheese varieties (via analysis of high-quality cheese samples and samples from cheeses with quality defects) remains an important step in making HT-qPCR even more useful for future diagnostic applications. The extended automation of DNA extraction could make the analysis even more cost-effective. This could facilitate the ultimate vision of using HT-qPCR in the future as a cost-effective and rapid method for monitoring the bacterial community composition, thereby assuring the quality of raw milk cheeses.

5.3 References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46**,** W537-W544. doi: 10.1093/nar/gky379.

Afshari, R., Pillidge, C.J., Read, E., Rochfort, S., Dias, D.A., Osborn, A.M., et al. (2020). New insights into cheddar cheese microbiota-metabolome relationships revealed by integrative analysis of multi-omics data. *Sci. Rep.* 10**,** 3164. doi: 10.1038/s41598-020-59617-9.

Agroscope (2020). *Analytik Leistungen Agroscope* [Online]. Available: https://www.agroscope.admin.ch/dam/agroscope/de/dokumente/services/analytik/lebensmittelanalytik/analytik-leistungen-liebefeld.pdf.download.pdf/2020_Preisliste_Analytik_L_d_200128.pdf [Accessed June 21 2021].

Ascone, P., Maurer, J., Haldemann, J., Irmler, S., Berthoud, H., Portmann, R., et al. (2017). Prevalence and diversity of histamine-forming *Lactobacillus parabuchneri* strains in raw milk and cheese A case study. *Int. Dairy J.* 70**,** 26-33. doi: 10.1016/j.idairyj.2016.11.012.

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

Bordin, N., and Devos, D.P. (2018). ICBdocker: a Docker image for proteome annotation and visualization. *Bioinformatics* 34**,** 3937-3938. doi: 10.1093/bioinformatics/bty493.

Burtscher, J., Hobl, L., Kneifel, W., and Domig, K.J. (2020). Short communication: Clostridial spore counts in vat milk of Alpine dairies. *J. Dairy Sci.* 103**,** 2111-2116. doi: 10.3168/jds.2019-17559.

Cocolin, L., Dolci, P., Rantsiou, K., Urso, R., Cantoni, C., and Comi, G. (2009). Lactic acid bacteria ecology of three traditional fermented sausages produced in the North of Italy as determined by molecular methods. *Meat Sci* 82**,** 125-132. doi: 10.1016/j.meatsci.2009.01.004.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326**,** 1694-1697. doi: 10.1126/science.1177486.

Desfosses-Foucault, E., Dussault-Lepage, V., Le Boucher, C., Savard, P., Lapointe, G., and Roy, D. (2012). Assessment of Probiotic Viability during Cheddar Cheese Manufacture and Ripening Using Propidium Monoazide-PCR Quantification. *Front. Microbiol.* 3**,** 350. doi: 10.3389/fmicb.2012.00350.

Desfosses-Foucault, E., LaPointe, G., and Roy, D. (2013). Dynamics and rRNA transcriptional activity of lactococci and lactobacilli during Cheddar cheese ripening. *Int. J. Food Microbiol.* 166**,** 117-124. doi: 10.1016/j.ijfoodmicro.2013.06.022.

Ganesan, B., Weimer, B.C., Pinzon, J., Dao Kong, N., Rompato, G., Brothersen, C., et al. (2014). Probiotic bacteria survive in Cheddar cheese and modify populations of other lactic acid bacteria. *J. Appl. Microbiol.* 116**,** 1642-1656. doi: 10.1111/jam.12482.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8**,** 2224. doi: 10.3389/fmicb.2017.02224.

Goecke, N.B., Krog, J.S., Hjulsager, C.K., Skovgaard, K., Harder, T.C., Breum, S.O., et al. (2018). Subtyping of Swine Influenza Viruses Using a High-Throughput Real-Time PCR Platform. *Front Cell Infect Microbiol* 8**,** 165. doi: 10.3389/fcimb.2018.00165.

Gómez-Torres, N., Garde, S., Peirotén, Á., and Ávila, M. (2015). Impact of Clostridium spp. on cheese characteristics: Microbiology, color, formation of volatile compounds and off-flavors. *Food Control* 56**,** 186-194. doi: 10.1016/j.foodcont.2015.03.025.

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

Jung, J.Y., Lee, S.H., Kim, J.M., Park, M.S., Bae, J.W., Hahn, Y., et al. (2011). Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl. Environ. Microbiol.* 77**,** 2264-2274. doi: 10.1128/AEM.02157-10.

Karunanithi, S., Simon, M., and Schulz, M.H. (2019). Automated analysis of small RNA datasets with RAPID. *PeerJ* 7**,** e6710. doi: 10.7717/peerj.6710.

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

Klijn, N., Nieuwenhof, F.F., Hoolwerf, J.D., van der Waals, C.B., and Weerkamp, A.H. (1995). Identification of *Clostridium tyrobutyricum* as the causative agent of late blowing in cheese by species-specific PCR amplification. *Appl Environ Microbiol* 61**,** 2919-2924.

LABGENE Scientific SA (2021). *4-Tube Strips & 4-Cap Strips For Rotor-Gene (REFERENCE: 3001002)* [Online]. Available: https://www.labgene.ch/pcr-strip-tubes/301-4-tube-strips-with-4-cap-strips-for-rotor-gene.html [Accessed June 21 2021].

Lauber, S., Herzog, F., Seidl, I., Böni, R., Bürgi, M., Gmür, P., et al. (2014). "Geht die Alprechnung auf?," in *Zukunft der Schweizer Alpwirtschaft. Fakten, Analysen und Denkanstösse aus dem Forschungsprogramm AlpFUTUR.,* ed. R. Wunderli. 80-93. Available: http://www.wsl.ch/ebooks/alpfutur/de.

Leite, M.F.A., and Kuramae, E.E. (2020). You must choose, but choose wisely: Model-based approaches for microbial community analysis. *Soil Biol. Biochem.* 151. doi: 10.1016/j.soilbio.2020.108042.

McGowan, T., Johnson, J.E., Kumar, P., Sajulga, R., Mehta, S., Jagtap, P.D., et al. (2020). Multi-omics Visualization Platform: An extensible Galaxy plug-in for multi-omics data visualization and exploration. *Gigascience* 9. doi: 10.1093/gigascience/giaa025.

Menegidio, F.B., Jabes, D.L., Costa de Oliveira, R., and Nunes, L.R. (2018). Dugong: a Docker image, based on Ubuntu Linux, focused on reproducibility and replicability for bioinformatics analyses. *Bioinformatics* 34**,** 514-515. doi: 10.1093/bioinformatics/btx554.

Panda, A., Chaudhari, N.M., and Tripathy, S. (2020). Genome Annotator Light (GAL): A Docker-based package for genome analysis and visualization. *Genomics* 112**,** 127-134. doi: 10.1016/j.ygeno.2019.03.012.

Pega, J., Rizzo, S., Perez, C.D., Rossetti, L., Diaz, G., Ruzal, S.M., et al. (2016). Effect of the addition of phytosterols and tocopherols on *Streptococcus thermophilus* robustness during industrial manufacture and ripening of a functional cheese as evaluated by qPCR and RT-qPCR. *Int. J. Food Microbiol.* 232**,** 117-125. doi: 10.1016/j.ijfoodmicro.2016.06.003.

Plengvidhya, V., Breidt, F., Jr., Lu, Z., and Fleming, H.P. (2007). DNA fingerprinting of lactic acid bacteria in sauerkraut fermentations. *Appl. Environ. Microbiol.* 73**,** 7697-7702. doi: 10.1128/AEM.01342-07.

Postollec, F., Falentin, H., Pavan, S., Combrisson, J., and Sohier, D. (2011). Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiol.* 28**,** 848-861. doi: 10.1016/j.fm.2011.02.008.

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

Rautiainen, M., and Marschall, T. (2020). GraphAligner: rapid and versatile sequence-to-graph alignment. *Genome Biol.* 21**,** 253. doi: 10.1186/s13059-020-02157-2.

Ritschard, J.S., Amato, L., Kumar, Y., Muller, B., Meile, L., and Schuppler, M. (2018). The role of the surface smear microbiome in the development of defective smear on surface-ripened red-smear cheese. *AIMS Microbiol* 4**,** 622-641. doi: 10.3934/microbiol.2018.4.622.

Ruggirello, M., Cocolin, L., and Dolci, P. (2016). Fate of *Lactococcus lactis* starter cultures during late ripening in cheese models. *Food Microbiol.* 59**,** 112-118. doi: 10.1016/j.fm.2016.05.001.

Ruggirello, M., Giordano, M., Bertolino, M., Ferrocino, I., Cocolin, L., and Dolci, P. (2018). Study of *Lactococcus lactis* during advanced ripening stages of model cheeses characterized by GC-MS. *Food Microbiol.* 74**,** 132-142. doi: 10.1016/j.fm.2018.03.012.

Singh, U., and Wurtele, E.S. (2021). orfipy: a fast and flexible tool for extracting ORFs. *Bioinformatics*. doi: 10.1093/bioinformatics/btab090.

Tamang, J.P., Watanabe, K., and Holzapfel, W.H. (2016). Review: Diversity of Microorganisms in Global Fermented Foods and Beverages. *Front. Microbiol.* 7**,** 377. doi: 10.3389/fmicb.2016.00377.

Turgay, M., Schaeren, W., Graber, H.U., Wagner, E., Amrein, R., Bütikofer, U., et al. (2018). A field study investigating the effectiveness of vat milk controls by qPCR for the prevention of undesired propionic acid fermentation in Sbrinz PDO cheese. *Int. Dairy J.* 77**,** 80-88. doi: 10.1016/j.idairyj.2017.06.008.

Wibowo, M.C., Yang, Z., Borry, M., Hubner, A., Huang, K.D., Tierney, B.T., et al. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature*. doi: 10.1038/s41586-021-03532-0.

Wüthrich, D., Berthoud, H., Wechsler, D., Eugster, E., Irmler, S., and Bruggmann, R. (2017). The histidine decarboxylase gene cluster of *Lactobacillus parabuchneri* was gained by horizontal gene transfer and is mobile within the species. *Front. Microbiol.* 8**,** 218. doi: 10.3389/fmicb.2017.00218.

Xue, Z., Brooks, J.T., Quart, Z., Stevens, E.T., Kable, M.E., Heidenreich, J., et al. (2021). Microbiota assessments for the identification and confirmation of slit defect-causing bacteria in milk and Cheddar cheese. *mSystems* 6**,** e01114-01120. doi: 10.1128/mSystems.01114-20.

Yang, C., You, L., Kwok, L.-Y., Jin, H., Peng, J., Zhao, Z., et al. (2021). Strain-level multiomics analysis reveals significant variation in cheeses from different regions. *Lwt* 151. doi: 10.1016/j.lwt.2021.112043.

Zhao, Y., Su, J.Q., Ye, J., Rensing, C., Tardif, S., Zhu, Y.G., et al. (2019). AsChip: a high-throughput qPCR chip for comprehensive profiling of genes linked to microbial cycling of arsenic. *Environ. Sci. Technol.* 53**,** 798-807. doi: 10.1021/acs.est.8b03798.

Annex

I Poster presentations

Poster 1: Poster presented at the Annual PhD students meeting of the Doctoral School of Organismal Biology, Neuchâtel, Switzerland (March 29, 2017).

Agroscope, Food Microbial Systems | 2017 (Annual PhD students meeting, Neuchâtel, 2017)

Characterisation of the microbial community of unpasteurized milk cheeses

Dreier, Matthias
Agroscope , CH-3003 Bern; www.agroscope.ch

Introduction

Raw milk cheeses are considered richer in flavor than cheeses made from pasteurized milk due to the beneficial impact of the microbial community in raw milk. However, the impact of starter and non starter lactic acid bacteria (NSLAB) on cheese quality in terms of flavour, texture and ripening stability is still poorly understood in Swiss cheese varieties. So far, mainly culture-dependent methods with a limited set of selective media were used to study microbial communities of Swiss raw milk cheeses. The use of next generation sequencing (NGS) technologies offers new possibilities to study the diversity and composition of the microbial community during cheese ripening. However, bioinformatics analysis of such data is still a challenging and time consuming task and does not allow a reliable quantification of individual microbial species. Therefore, the present work aims to develop a new high-throughput qPCR approach for the rapid and cost-efficient quantification of microbial species in cheese. The use of this approach is expected to yield more comprehensive data on the growth and succession of microbial species during cheese ripening and thus will contribute to a better understanding of the impact of individual species on cheese quality. Moreover, high-throughput qPCR analysis of cheeses will offer new diagnostic possibilities regarding the rapid identification of microbial induced cheese quality defects.

Approach

- Identify relevant bacteria important for cheese quality.

Species	Category
Lactococcus lactis	Starter
Streptococcus thermophilus	Starter
Lactobacillus helveticus	Starter
Lactobacillus delbrueckii	Starter
Lactobacillus fermentum	Starter (whey)
Lactobacillus rhamnosus	NSLAB
Lactobacillus plantarum	NSLAB
Lactobacillus casei	NSLAB
Lactobacillus curvatus	NSLAB / BGA
Lactobacillus parabuchneri	NSLAB / BGA
Lactobacillus brevis	NSLAB / BGA
Enterococcus faecalis	NSLAB / BGA
Enterococcus faecium	NSLAB / BGA
Pediococcus acidilactici	NSLAB
Pediococcus pentosaceus	NSLAB
Propionibacterium freudenreichii	Adjunct
Leuconostoc mesenteroides	Adjunct
Clostridium tyrobutyricum	Butyric acid
BGA: biogenic amines	

- Development of individual qPCR primer pairs using a single set of amplification conditions.
- Check specificity and sensitivity of the primer pairs.
- Transfer to microfluidic high-throughput qPCR system.
- Validate high-throughput qPCR method with experimental cheeses made from micro-filtered milk with added target bacteria.
- Investigation of microbial community composition in cheese varieties made from unpasteurized milk.
- Analysis of commercial cheeses with quality defects.

Frequently observed cheese quality defects

- Excess gas formation leads to splits and cracks.
- Formation of histamine leads to a burning taste and can trigger allergic reactions in sensitive individuals
- Flavour and discoloration defects lead to consumers refusal of the product.
- Faulty fermentation such as propionic or butyric acid fermentation

High-throughput qPCR

advantages:

- quantitative, fast and reliable
- no laborious bioinformatics

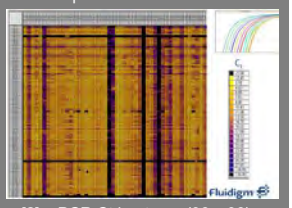
drawbacks:

- only detection of defined targets
- new qPCR assays working with identical amplification conditions are required

Microfluidic high-throughput qPCR (Fluidigm)

- Dynamic array chip: 48 samples and 48 qPCR assays
- Each sample is mixed with each assay by an integrated fluidic controller (IFC).
- Performs 2304 qPCR reactions in parallel

[1]: 48.48 dynamic array chip



[2]: qPCR Ct heatmap (96 x 96)

Fluidigm, 2015, BioMark™ HD Data Collection user guide

Agroscope good food, healthy environment

UNIVERSITÉ DE NEUCHÂTEL

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
Agroscope

Poster 2: Poster presented at the Agroscope PhD/Postdoc Symposium, Agroscope Reckenholz, Zürich, Switzerland (September 13, 2018).

Agroscope | 2018

SpeciesPrimer: A bioinformatics pipeline dedicated to the design of qPCR primers for the quantification of bacterial species

Matthias Dreier[1, 2], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1], and Pilar Junier[2]
[1] Agroscope, Schwarzenburgstrasse 161, CH-3003 Bern, Switzerland
[2] University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

Introduction

We aim to develop a high-throughput qPCR approach for the rapid and cost-effective quantification of microbial species in cheese. The main challenges to the successful development of qPCR assays are the identification of target-specific nucleotide sequences and the design of specific primers. Today, most available tools for primer design require either laborious manual manipulations or high-performance computing systems. Faster tools have been developed only recently, offering the possibility to build high-throughput pipelines for automated primer design on a standard desktop computer. We developed a pipeline, named SpeciesPrimer, for automated high-throughput screening for species-specific target regions and the design of dedicated primers.

Methods

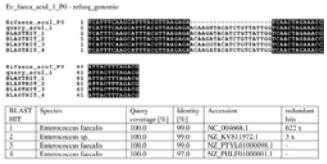
SpeciesPrimer pipeline

- Primer design (default settings)
- Optimization of input and settings

Pipeline workflow	Tools
Download and annotation of genome assemblies	NCBI Entrez Utilities Biopython Prokka
Genome assembly quality control	BLAST
Pan-genome analysis	Roary
Core gene phylogeny	FastTree2
Processing of core gene sequences	SQlite3 Prank Consambig (EMBOSS) GNU parallel
BLAST of conserved, single copy sequences	BLAST
Primer design for species specific sequences	Primer3
Primer quality control	BLAST Mfold MFEPrimer 2.0

***In-silico* validation**

- Web-based BLAST of the PCR product (DB: genomes_refseq)
- Primer-BLAST of primer pairs (DB: refseq_representative_genomes)

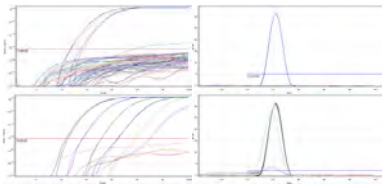


Summarized alignments of a web-based BLAST search with the predicted PCR product of the primer pair as query.

***In-vitro* validation**

qPCR with DNA from pure cultures:

- Inclusivity: 20 – 25 strains (target)
- Exclusivity: 122 strains (non-target)
- Dynamic range and efficiency: Dilution series of type strain genomic DNA (10^6 – 10 copies / reaction)



Amplification curves and melting curve analysis for the qPCR assay Ecfaeca_acuI. Exclusivity validation (upper) and dilution series (lower).

Results

SpeciesPrimer input

- List of non-target species names (abundant in cheese)
- NCBI pre-formatted nucleotide collection database (nt)
- Species names:

	# Genomes
Enterococcus faecalis	390
Enterococcus faecium	575
Pediococcus acidilactici	127
Pediococcus pentosaceus	19

SpeciesPrimer output

Species	# Primer
Enterococcus faecalis	12 / -
Enterococcus faecium	1 / *8
Pediococcus acidilactici	2 / *14
Pediococcus pentosaceus	58 / *19

** Primer systems after optimization of the SpeciesPrimer input (genomes) and parameters*

Target species	Primer pair	PPC	Primer match	BLAST	primer-BLAST
E. faecalis	Ecfaeca_acuI	100	625/625	specific	specific
	Ecfaeca_lacF	94.	624/625	specific	specific
	Ecfaeca_g2706_1	90.6	622/625	specific	specific
	Ecfaeca_dnaE	97.	624/625	specific	specific
E. faecium	Ecfaeci_g2948_1	96.9	908/914	specific	specific
	Ecfaeci_g2948_2	100	908/914	specific	specific
	Ecfaeci_g2948_0	100	906/913	specific	specific
	Ecfaeci_rsmB	96.6	908/915	specific	specific
P. acidilactici	Pdacidi_g1727	100	22/22	specific	specific
	Pdacidi_g5404	96.6	22/22	specific	specific
	Pdacidi_g5571	100	22/22	specific	specific
	Pdacidi_g1740_4	92.6	22/22	specific	specific
P. pentosaceus	Pdpento_tagE	100	12/12	specific	specific
	Pdpento_tagB	96.6	12/12	specific	specific
	Pdpento_prmC	100	12/12	specific	specific
	Pdpento_yycI	96.6	12/12	specific	specific

Primer candidates with a high specificity for the target genes were selected.

PPC: Primer pair coverage score (MFEprimer 2.0).
Primer match: Aligned sequences with perfect match (primer binding region) / total number of aligned sequences
Specific: BLAST hits only for target species genome sequences.

	E. faecalis	*P. acidilactici*			
Target gene	*acuI*	*asnS*	*g1164*	**g1727*	**g5571*
Inclusivity	22 / 22	21 / 21	21 / 21	20 / 21	21 / 21
Exclusivity	0 / 122	0 / 122	6 / 122	31 / 122	22 / 122
Efficiency	100 %	106 %	91 %	102 %	97 %

	E. faecium			*P. pentosaceus*	
Target gene	*purD*	**g2948*	**rsmB*	**tagE*	**prmC*
Inclusivity	24 / 25	25 / 25	25 / 25	*27 / 27*	*27 / 27*
Exclusivity	0 / 122	0 / 122	4 / 122	*18 / 122*	*2 / 122*
Efficiency	99 %	92 %	101 %	*100 %*	*100 %*

Specific qPCR assays for *E. faecalis (acuI), E. faecium (g2948), P. acidilactici (asnS)* and *P. pentosaceus (prmC)* were identified. The *in-vitro* validation of these qPCR assays showed satisfying inclusivity, exclusivity and amplification efficiencies.

Summary

- The SpeciesPrimer pipeline and a validation work-flow were used to design and validate qPCR assays
- Four qPCR assays with good performance under the exact same PCR cycling conditions were developed
- SpeciesPrimer is available for download at https://hub.docker.com/r/biologger/speciesprimer

Agroscope good food, healthy environment

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
Agroscope

Development of a high-throughput qPCR system for the quantitative investigation of bacterial communities in cheese

Matthias Dreier[1, 2], Hélène Berthoud[1], Noam Shani[1], Daniel Wechsler[1], and Pilar Junier[2]

[1] Agroscope, Schwarzenburgstrasse 161, CH-3003 Bern, Switzerland

[2] University of Neuchâtel, Rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

Background

The composition of the cheese microbiome has an important impact on the sensorial quality and safety of cheese. Therefore, much effort has been made to investigate the cheese microbial community composition. Quantitative real-time PCR (qPCR) is a well-established method for detecting and quantifying bacteria, and its quantitative aspects advantageously complement 16S rRNA gene amplicon sequencing in determining the microbial community composition. High-throughput qPCR (HT-qPCR) using microfluidics brings further advantages by providing fast results and by decreasing the cost per sample. The aim of this study was to develop and validate a rapid and cost efficient HT-qPCR system for the quantitative characterization of bacteria relevant to cheese quality.

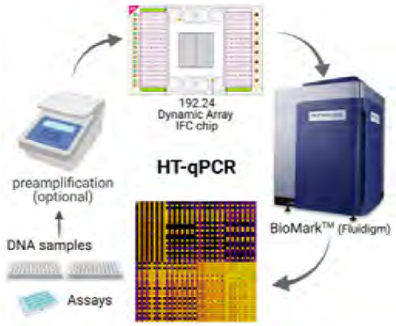


Figure 1: Overview of the HT-qPCR workflow[1, 2]

Methodology

- **Primer design**
  - **24 species-specific primers** with a single set of amplification conditions[3]
- **Evaluation**
  - **Inclusivity and exclusivity** of the qPCR assays were evaluated using DNA from single strains of the target species and off-target type strains.
  - **Application on cheese DNA samples**
    Raclette-type model cheeses inoculated with different target species

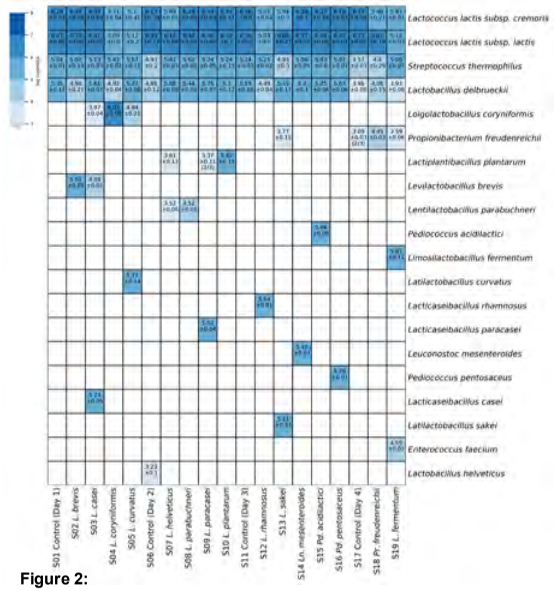


Figure 2:
HT-qPCR results of Raclette-type model cheese samples produced with selected adjunct species.

Results

- **Highly specific HT-qPCR assays** for 24 bacterial species relevant for cheese quality.
- **Quantification of inoculated target species** in Raclette-type model cheese DNA samples.
- **Species relevant to the organoleptic quality of cheese were quantified, despite a high detection limit** (800 copies/µL) due to the nanoliter-scale reactions used in the microfluidic qPCR system.
- **Preamplification** could be used to detect bacterial species with low abundance (< 8×10^4 genome equivalents/g cheese).

Summary

- **Fast, accurate, and cost-efficient** monitoring of desired and undesired microorganisms in cheese.
- **Simultaneous screening of 24 species** in 56 cheese DNA samples (technical triplicates).
- The application of HT-qPCR could be very useful for monitoring the composition of the ripening microbiota, thereby ensuring a constant product quality in the production of raw milk cheese.

References

[1] Created with BioRender (https://www.biorender.com)

[2] https://www.fluidigm.com/

[3] https://github.com/biologger/speciesprimer

Publication available at
https://doi.org/10.3389/fmicb.2020.619166

UNIVERSITÉ DE NEUCHÂTEL

GDC Genetic Diversity Centre Zurich

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
Agroscope

II Curriculum Vitae

Matthias Dreier

Sonnenweg 4a
3098 Köniz

Email: matthias.dreier@agroscope.admin.ch

Born: March 5, 1987
Nationality: Swiss

Future position

2021-2024 **Postdoctoral researcher**
Research group Cheese Quality, Cultures and Terroir, Agroscope

Education

2016-present **PhD in Microbiology**
Place: Laboratory of Microbiology, Institute of Biology, University of Neuchatel
Research group Cheese Quality, Cultures and Terroir, Agroscope
Thesis: Characterization of the microbial community in raw milk cheeses by high-throughput qPCR
Supervisor: Prof. Pilar Junier

2011-2013 **Master of Science in Molecular Life Sciences**
Place: Nonsense group, Department of Chemistry and Biochemistry, University of Bern
Thesis: Generation of stable cell lines for investigation of nuclear nonsense-mediated mRNA decay (NMD) effects by live-cell imaging
Supervisor: Prof. Oliver Mühlemann

2007-2011 **Bachelor of Science in Biochemistry**
Place: Nonsense group, Department of Chemistry and Biochemistry, University of Bern
Thesis: Half-life determination of different NMD reporter mRNAs
Supervisor: Prof. Oliver Mühlemann

Professional experience

2015-2016 Research associate
Place Research group cheese quality, Institute for food science, Agroscope

2015 Civilian service
Place Institute for food science, Agroscope

Publications

2022 Dreier M, Meola M, Berthoud H, Shani N, Wechsler D, Junier P. High-throughput qPCR and 16S rRNA gene amplicon sequencing as complementary methods for the investigation of the cheese microbiota. BMC Microbiol. 2022;22(1); doi: 10.1186/s12866-022-02451-y.

2021 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; doi: 10.3389/fmicb.2020.619166.

Dreier M, Wechsler D. High-throughput quantification of cheese bacteria. Chimia (Aarau). 2021;75(6):550; doi: 10.2533/chimia.2021.550.

2020 Burtscher J, Kuller F, Dreier M, Arias-Roth E, Drissner D, Domig KJ. Characterization of *Clostridium tyrobutyricum* strains using three different typing techniques. Microorganisms. 2020;8(7); doi: 10.3390/microorganisms8071057.

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; doi: 10.7717/peerj.8544.